

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 30, 2002, 00:38:12 ; Search time 1750.92 Seconds
(without alignments)
5858.455 Million cell updates/sec

Title: us-09-821-821-1

Perfect score: 760

Sequence: 1 ttccagtgctccagcagcc.....aaaaaaaaaaaaaaaaaaaaa 760

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

- 1: em_estba.*
- 2: em_esthum.*
- 3: em_estin.*
- 4: em_estmu.*
- 5: em_estov.*
- 6: em_estpl.*
- 7: em_estro.*
- 8: em_htc.*
- 9: gb_est1.*
- 10: gb_est2.*
- 11: gb_htc.*
- 12: gb_gss.*
- 13: em_gss_hum.*
- 14: em_gss_inv.*
- 15: em_gss_pln.*
- 16: em_gss_vrt.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	707.2	93.1	713	10	BI829760
2	504.8	66.4	516	9	AA436088
3	503.2	66.2	508	9	AI149899
4	382.2	50.3	387	9	AA416972
5	377.2	49.6	382	9	AA435988
6	325.4	42.8	394	9	AA435983
7	300.4	39.5	415	9	AA470059
8	295.2	38.8	484	9	BE107659
9	271	35.7	482	9	AI002083
10	267.4	35.2	410	9	AA411806
11	245.4	32.3	389	9	AA781801
12	148.4	19.5	538	12	AQ108532
13	145	19.1	281	9	AA707529
14	144.2	19.0	332	10	BF319786
15	135	17.8	601	10	BE638317
16	124.6	16.4	224	10	BE638325
17	95	12.5	568	10	BM253528

18	86.4	11.4	790	10	BI771845
19	86.4	11.4	1071	9	AL544561
20	85	11.2	900	10	BI759693
21	84	11.1	793	10	BG484817
22	82.8	10.9	799	10	BI908709
23	82.2	10.8	704	10	BG571626
24	82	10.8	857	10	BG720182
25	81.4	10.7	218	10	BE638321
26	80.4	10.6	508	9	AA234138
27	80.4	10.6	562	9	AA418443
28	79.2	10.4	517	10	BG664654
29	76.2	10.0	713	10	BI562282
30	72.2	9.5	421	12	B86842
31	70.8	9.3	899	10	BG720623
32	68.6	9.0	921	10	BI561611
33	68.2	8.9	856	10	BI560592
34	67.6	8.9	775	10	BI560610
35	66.8	8.8	689	10	BG719295
36	59.6	7.8	887	9	AL531049
37	59.4	7.8	715	10	BG538851
38	58.8	7.7	579	10	BE513276
39	58.6	7.7	470	9	AI807884
40	58.2	7.7	528	10	BE675149
41	58.2	7.7	551	10	BM152988
42	58.2	7.7	904	10	BG024663
43	58	7.6	770	10	BI464864
44	57.2	7.5	947	11	AK008652
45	56	7.4	1504	11	AK003110

ALIGNMENTS

RESULT 1

BI829760	BI829760	713 bp	mRNA	linear	EST 04-OCT-2001
LOCUS	603079831F1 NIH_MGC_119	Homo sapiens	CDNA clone	IMAGE:5171526 5',	
DEFINITION	mRNA sequence.				
ACCESSION	BI829760				
VERSION	BI829760.1	GI:15941310			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	1 (bases 1 to 713)				
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	Unpublished (1999)				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs@mail.nih.gov Tissue Procurement: Life Technologies, Inc. CDNA Library Preparation: Life Technologies, Inc. CDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM1426 row: n column: 07 High quality sequence stop: 713.				

FEATURES source

1. .713
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5171526"
/clone_lib="NIH_MGC_119"
/tissue_type="medulla"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMV-SPORT6; Site: 1; NotI: anonymous male age 27. Library is oligo-dr primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.3 kb, insert size range

0.9-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 013. Note: this is a NIH_MGC Library."

BASE COUNT 202 a 154 c 123 g 233 t 1 others

ORIGIN

Query Match 93.1%; Score 707.2; DB 10; Length 713;
Best Local Similarity 99.4%; Pred. No. 9.7e-90;
Matches 709; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 12 caggcagctcagcacaagaaagacatggtctagactgaagtaccactaaatcatct 71
|||||
DB 1 CAGGAGCCCTCAGCACAGAAAGAACATGGTCTAGACTGAAGTACCACTAAATCATCT 60
|||||

QY 72 cctttcaaaattaccgcagcaccatcatgattcaagcaccgcacacagtcggtgtttc 131
|||||
DB 61 CCTTCAAAATTATCAGGACACCATCATGGATTCAGACCCGCACACAGTCGGGTGTTTC 120
|||||

QY 132 tggatttctccagaataactgcttcagaatgatggtccacagaactttcagccacga 191
|||||
DB 121 TGGTATTTCCTCCAGAAATCAGTCTTCAGAAATGATGATGATGATGATGATGATGATGAT 180
|||||

QY 192 cctttcaactcaagcccttcgaaattatttctgagaaataaataaatacttaggga 251
|||||
DB 181 CCTTTCACTCAAGCCCTTGCAAAATATTGCTGAAAAATGAAATCTTAGGGA 240
|||||

QY 252 ctatccagatcctgttgggaattatgacctttcttttggagttatctctctttcaactt 311
|||||
DB 241 CTATCCAGATCCGTGTTGGAAATATGACCTTTCCTTTGGAGTTATCTCTCTTTTACCT 300
|||||

QY 312 tgttaaacacatccaaggtttccctttatattttcttcagagatccattctggggt 371
|||||
DB 301 TGTAAAAACCATATCCAAGCTTTCCCTTTATATTCTTTTCAGGATATCCATCTCGGGCT 360
|||||

QY 372 ctgtttgttcattaatcttgagccttcctaattcagtgagaaagaaacacagaaa 431
|||||
DB 361 CTGTTTNGTTTCAATTAATCTGAGCCCTTCCTAATTCAGTGAAGAAAGAAACACAGAAA 420
|||||

QY 432 ctctgataatattgagccgaataatgaattttctagtcctctgggagcaatagctggaa 491
|||||
DB 421 CTCTGATAATATTGAGCGGAATTAATGAATTTCTTAGTCCCTGGGAGCAATAGCTGGAA 480
|||||

QY 492 tcattctctccattggtttcattcctcagatcaaaactacatttgggttatctcacc 551
|||||
DB 481 TCATTCTCTCACATTTGGTTTCATCCTAGATCAAAACTACATTTGTGGTTATCTCACC 540
|||||

QY 552 aaaaatgctagtgtaagctgttactgtctgttcttgggaatttttgattacattgatga 611
|||||
DB 541 AAATAGTCAGTGTAGGCTGTACTGTCTCTGTCTGTGGGAATTTTGATACATTTGATGA 600
|||||

QY 612 ctttcagcattattgaattattcattctctgctttctcaattttgggggtgacctcag 671
|||||
DB 601 CTTTCACCATATTGAATTTATTCATTTCTGCTTTCATTTTCGGGTGCCACTCAG 660
|||||

QY 672 agattgtgattgtaacaatgttgttgactagcactgtgagaaataaagatgt 724
|||||
DB 661 AGGATTGTGTTGTGAACAATGTCGTTGACTAGTACTGTGAGAAATAAAGATGT 713
|||||

RESULT 2

AA436088
LOCUS z403a08.r1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:730742
DEFINITION 5' similar to SW:CD20_HUMAN P11836 B-LYMPHOCYTE ANTIGEN CD20 ;,
mRNA sequence.

ACCESSION AA436088
VERSION AA436088.1 GI:2141002
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 516)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Gelsel,G., Jost,S., Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theisling,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE WashU-NCI human EST Project
JOURNAL Unpublished (1997)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 733 Std Error: 0.00
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 482.

FEATURES
source location/Qualifiers
1..516
/organism="Homo sapiens"
/db_xref="GDB:5927437"
/db_xref="taxon:9606"
/clone="IMAGE:730742"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGGGCGGCCCAATTTTTTTTTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p7T3 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 149 a 117 c 80 g 170 t

Query Match 66.4%; Score 504.8; DB 9; Length 516;
Best Local Similarity 99.6%; Pred. No. 1.7e-61;
Matches 506; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 47 gactgaagtaccacaataatcatctcctttcaaatattaccacacacacatgatgttca 106
|||||
DB 1 GACTGAAGTACCAACTAAGTCACTCTCTTCAAAATTATCACCAGACCATCATCGATTCA 60
|||||

QY 107 agcaccgcacagtcgcggtgtttctggtatttctccacagaatacactgcttcagaatat 166
|||||
DB 61 AGCACCAGCACAGTCCGGTGTCTTCTGTTATTTCTCTCAGAAATCACTGCTTCAGAAAT 120
|||||

QY 167 gagtccacagaacttccagcagacctttcaactcaaaagcccttgcagaaattattt 226
|||||
DB 121 GAGTCCACAGAACTTTCAGCCAGCACCTTTTCAACTCAAGCCCTTTCGCAAAATTTAT 180
|||||

QY 227 gctagaaaaatgaaatcttagggactatccagatccctgtttggaattatgacctttct 286
|||||
DB 181 GCTAGAAAAATGAAATCTTAGGGACTATCCAGATCTCTGTTGGAAATATGACCTTTCT 240
|||||

QY 287 ttggagttattctctcttttcaactttgttaaaacatatccagggttccctttatattt 346
|||||
DB 241 TTTGGAGTTATCTTCTCTTTTTCACCTTTGTTAAACCATATCCAAAGGTTTCCCTTTATAT 300
|||||

QY 347 ctttcagagatatccattctggggtctgttttctggttcatttaattcttgagccttctaatt 406
|||||
DB 301 CTTTCAGGATATCCATTCTGGGGCTCTGTTTGTTCATTAATTCTGGAGCCTTCTTAAT 360
|||||

QY 407 gcagtgaagaagaaaaaacacagaaactctgataattatgagccgaataaagaaatttctt 466
|||||
DB 361 GCAGTGAAGAAAGAAACACAGAAACTCTGATATATTATGACCCGAATAATGAAATTTCTT 420
|||||

QY	350	tcagatataccattctgggctctgtttttgttcattcaattctggagccttcctaattgca	409
Db	388	TCAGGATATCCCAATCTGGGCTCTGTTTGTTCATTAATTTCTGGAGCCCTCTCTAATGCA	329
QY	410	gtgaaagaaaaaacacagaaactctgataatattggagcgaataatgaatttcttagt	469
Db	328	GTGAAAGAAAAACACAGAAACTCTGATATATATTGAGCCGAATAATGAATTTCTTAGT	269
QY	470	gccctgggagcaatagctgggaatccttctctcacatttggtttctcctagatcaaac	529
Db	268	GCCTGGGAGCAATAGCTGGAAATCATTTCTCTACATTTGGTTTCATCTAGATCAAAAC	209
QY	530	tacatttgggtattctccacaaaatagtcagtgtaaggctgttactgtctctgttcttg	589
Db	208	TACATTTGTGGTATATCTCCACAAAATAGTCAGTGTGAAGCTGTCTACTGCTGTGTTCTTG	149
QY	590	ggaattttgatcacattgatgacttccagcattattgaattattcatttctctgccttc	649
Db	148	GGAAATTTTTCATTGATGACTTTTCAGCATTTATTCATTTCTCTGCCTTC	89
QY	650	tcaatttgggtggcactcagagattgtgattgtgaaacatgttggactagtagcactg	709
Db	88	TCAATTTTGGGGTGGCACTCAGAGATTTGATTTGAAACAATGTTGTGACTAGCAGCTG	29
QY	710	tgagaataaagatgtgttaaaactctcaa	737
Db	28	TGAGATAAAGATGTGTTAAATAAAAA	1
RESULT	4		
AA416972/c			
LOCUS	AA416972	387 bp	linear
DEFINITION	zt94h05.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:730041		EST 09-NOV-1997
	3' similar to SW:CD20_HUMAN P11836 B-LYMPHOCYTE ANTIGEN CD20 ;		
	nrna sequence.		
ACCESSION	AA416972		
VERSION	AA416972.1	GI:2077080	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 387)		
	Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisels,G., Jost,S.,		
	Krizman,D., Kucaba,T., Lacy,M., Le.N., Lennon,G., Marra,M., Martin		
	J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,		
	White,Y., Wyllie,T., Waterston,R. and Willson,R.		
TITLE	WASP-NCI human EST Project		
JOURNAL	Unpublished (1997)		
COMMENT	Contact: Wilson RK		
	Washington University School of Medicine		
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108		
	Tel: 314 286 1800		
	Fax: 314 286 1810		
	Email: est@wustl.wustl.edu		
	This clone is available royalty-free through LML; contact the		
	IMAGE Consortium (info@image.llnl.gov) for further information.		
	Insert Length: 656 Std Error: 0.00		
	Seq primer: -41m13 fwd,ET from Amersham.		
FEATURES	Location/Qualifiers		
source	1. 387		
	/organism="Homo sapiens"		
	/db_xref="GDB:5926570"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:730041"		
	/clone.lib="Soares_testis_NHT"		
	/sex="male"		
	/lab_host="DH10B"		
	/note=Vector: pWT3D-Pac (Pharmacia) with a modified		
	polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA		
	was prepared from mRNA obtained from Clontech Laboratories		
	, Inc., and primed with a Not I - oligo(dT) primer [5].		

Seq primer: -41ml3 fwd, ET from Amersham
High quality sequence stop: 280.
Location/Qualifiers
I. .382
/organism="Homo sapiens"
/db_xref="GDB:5927437"
/db_xref="taxon:9606"
/clone_lib="IMAGE:730742"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - oligo(dt) primer [5',
TGTTACCAATCTGAATGGGCGCGCCCAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT
ORIGIN

138 a 77 c 67 g 105 t

Query Match 50.3%; Score 382.2; DB 9; Length 387;
Best Local Similarity 99.2%; Pred. No. 2.4e-44;
Matches 384; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 346 tcttcaggatcattctgggctctgtttgttcatttaatttgagccttcctaat 405
|||||
Db 387 TCTTAGGGATATCCATCTCGGGCTCTGTTTGTTCATTAATCTGAGCCTTCTTAAT 328
QY 406 tgcagtgaaagaaacacagaaactctgataatattaggccgaataagaattttct 465
|||||
Db 327 TGCAGTGAAGAAAGAAACACAGAACTCTGATAATATTGAGCCGAATAATGAATTTC 268
QY 466 tagtgcctggagcaatagctgaatcattctctccattgtttcattcctcctagatca 525
|||||
Db 267 TAGTGCCTGGGAGCAATAGCTGAATCATCTCTCACAATTTGTTTCATCTTAGATCA 208
QY 526 aaactacattgtgttattctcaccataatagctcagtgataagctgttactgtcctgt 585
|||||
Db 207 AACACATATTGTGGTATTCTCCCAAAATGCTAGTGAGGCTGTACTGCTCCTGTT 148
QY 586 ctggggaatttgattacattgactcttcagcattattggaattattcattctctgcc 645
Db 147 CTGGAATTTTGAATTCATATGATGACTTTCAGCAATATTGAATTTATTCATTTCTCTGCC 88
QY 646 ttctcatttgggtggccactcagagatgtgaattgtgaacaattgttgactagc 705
Db 87 TTTCTCAATTTGGGGTGGCCACTCAGAGGATGTGATGTGAACAATGTTGTTGACTAGC 28
QY 706 actgtgagaataaagatgtgttaaat 732
Db 27 ACTGTGAGAAATAAGATGTGTTAAAT 1

RESULT 5
AA435988/c
LOCUS
DEFINITION

AA435988 382 bp mRNA linear EST 09-NOV-1997
zu03a08.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:730742
3' similar to SW:CD20_HUMAN P11836 B-LYMPHOCYTE ANTIGEN CD20 ;,
mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE

AA435988.1 GI:2140902
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 382)
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin
J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,
White, Y., Wylie, T., Waterston, R. and Wilson, R.
WashU-NCI human EST Project
Unpublished (1997)
Contact: Wilton RK

TITLE
JOURNAL
COMMENT

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estevaton.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone; similarity on wrong strand
Insert length: 733 Std Error: 0.00

FEATURES
source

/organism="Homo sapiens"
/db_xref="GDB:5927437"
/db_xref="taxon:9606"
/clone_lib="IMAGE:730742"
/sex="male"
/lab_host="DH10B"

BASE COUNT 135 a 73 c 67 g 107 t
ORIGIN

Query Match 49.6%; Score 377.2; DB 9; Length 382;
Best Local Similarity 99.2%; Pred. No. 1.2e-43;
Matches 379; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 356 tatccattctgggctctgtttgttcatttaattctggagccttcctaatgcagtgaaa 415
Db 382 TATCCATCTCGGGCTCTGTTTGTTCATTAATTTGAGCCCTCCTTAATTTGCAAGTAAA 323
QY 416 aaaaaaacacagaaactctgataatattgagccgaataaataattcttagtgcctg 475
Db 322 AGAAAAACACAGAACTCTGATAATATTGAGCCGAATAATGAATTTCTTAGTGCCTG 263
QY 476 ggagcaatagctgggaatcattctcctcattgtgttcattcctagatcaaaactacatt 535
Db 262 GGAGCAATAGCTGGAATCATTTCTCCACATTTGGTTTTCATCTCCTAGATCAAACTCAT 203
QY 536 tgtgttattctcaccacaaatagctcagtgtaagctgttactgcctgttcttgggaatt 595
Db 202 TGTGTTATTCTCACCAAAATAGTCAGTCTAAGGCTGTACTGTCTCTTCTTGGGAAT 143
QY 596 ttgattacattgagcttccagcattattgataattcattctctgcctcttctcaatt 655
Db 142 TTGATTACATTTGATGACTTTTCAGCATTTATTGAATTTATTCATTTCTGCTGCTCAATT 83
QY 656 ttgggtgcccactcagaggattgtgatttgaaacattgtgtgactagcactgtgagaa 715
Db 82 TTGGGTGCCACTCAGAGGATTGTGATTGTGAACAATGTTGTTGACTAGCACTGTGAGAA 23
QY 716 taaagatgtgttaaaatctcaa 737
Db 22 TAAAGATGTGTTAAATATATAA 1

RESULT 6
AA758635/c
LOCUS
DEFINITION

AA758635 394 bp mRNA linear EST 29-DEC-1998
ab67b04.s1 Soares_testis_NHT Homo sapiens cDNA clone 1320655 3',
similar to SW:CD20_HUMAN P11836 B-LYMPHOCYTE ANTIGEN CD20 ;,
sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE

AA758635.1 GI:2806498
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 394)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapb@rmail.nih.gov
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
 , Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the i.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bbrp/image/image.html
 Insert Length: 661 Std Error: 0.00
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 390.

FEATURES
 source
 1..394

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="1320655"
 /clone_lib="Soares_testis_NHT"
 /sex="male"
 /lab_host="DH10B"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
 polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
 was prepared from mRNA obtained from Clontech Laboratories
 , Inc., and primed with a Not I - oligo(dT) primer [5',
 TGTACCAATCTGAAGTGGAGCGCGCCCAATTTTTTTTTTTT 3'].
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT7T3 vector. Library
 went through one round of normalization to Cot5, and was
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 133 a 73 c 67 g 121 t

ORIGIN

Query Match 42.8%; Score 325.4; DB 9; Length 394;
 Best Local Similarity 96.2%; Pred. No. 1.9e-36;
 Matches 376; Conservative 0; Mismatches 11; Indels 4; Gaps 4;

QY 346 tcttcaggatattccattgggctctgtttgttcatttaattctgagccttccctaa 405

|||||

Db 394 TCTTTCAGGATATCCATCTCGGGCTCTGTATGATATCATTAATTCGGAGCCTTCCTAAT 335

|||||

QY 406 tgcagtgaaagaaaaaacacagaaactctgataatattgagccgaataaattttct 465

|||||

Db 334 TGCAGTGAAGAAAGAAAACACAGAAACTCTGATAATATTGAGCCGAATAATGAATTTCT 275

|||||

QY 466 tagtgccttgaggagcaatagctggaatcatctccacatttg-gtttcaccttagatc 524

|||||

Db 274 TAGTGCCCTGGAGCAATAGCTGGATCATCTTCATCATCATTTTGAGTTTCATCTAGATC 215

|||||

QY 525 aaactacatttggttattctcaccaaaatagtcagtgtaaggctgttaactgtcctgt 584

|||||

Db 214 AAAACTACATTGTGTGATTATCTCACAAATAGTCAGTGAAGCTGTACTGTCTCTGA 155

|||||

QY 585 tcttggaattttgatcacattgatgactttcagcat-tattgaattattcat-ttctct 642

|||||

Db 154 TCTTGGGAATTAGATACATTGACITTCAGCATATATTGAATTATTCATATTCCT 95

|||||

QY 643 gccctt-ctcaattttggggtgccactcagaggattgtgattgtaacaaattgtttgac 701

|||||

Db 94 GCGTTTACTCAATTTTGGGTGGCCACTCAGAGGATTGTGATTGTGAACAATGTAGTTGAC 35

|||||

QY 702 tagcacttgagaataaagatggttaaaat 732

|||||

Db 34 TAGCACTGTGAGATATGAATATGTTTAAT 4

|||||

RESULT
 AA470059
 LOCUS

AA470059 415 bp mRNA linear EST 09-NOV-1997

DEFINITION

zt94h05.r1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:730041
 5', similar to SW:CD20_HUMAN P11836 B-LYMPHOCYTE ANTIGEN CD20 ; ,
 mRNA sequence.

ACCESSION
 AA470059

VERSION
 AA470059.1 GI:2197368

KEYWORDS
 EST.

SOURCE
 human.

ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 415)
 Hillier,V., Allen,M., Bowles,L., Dubuque,T., Geisels,G., Jost,S.,
 Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
 ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
 White,Y., Wyllie,T., Waterston,R. and Wilson,R.

WashU-NCI human EST Project

Unpublished (1997)

Contact: Willson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert length: 656 Std Error: 0.00

Seq primer: -28m13 rev2 ET from Amersham

High quality sequence stop: 393.

FEATURES

source

1..415

/organism="Homo sapiens"

/db_xref="GDB:5926570"

/db_xref="taxon:9606"

/clone="IMAGE:730041"

/clone_lib="Soares_testis_NHT"

/sex="male"

/lab_host="DH10B"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
 polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
 was prepared from mRNA obtained from Clontech Laboratories
 , Inc., and primed with a Not I - oligo(dT) primer [5',
 TGTACCAATCTGAAGTGGAGCGCGCCCAATTTTTTTTTTTT 3'].

Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT7T3 vector. Library
 went through one round of normalization to Cot5, and was
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 129 a 96 c 67 g 123 t

ORIGIN

Query Match 39.5%; Score 300.4; DB 9; Length 415;

Best Local Similarity 79.9%; Pred. No. 5.5e-33;

Matches 414; Conservative 0; Mismatches 1; Indels 103; Gaps 1;

QY 42 gctagactgaagtaccacaaatcatctctcttcaaatatcacccagaccatcattg 101

|||||

Db 1 GACTAGACTGAAGTACCACAACTAAATCATCTCTCTTCAATATATCACCAGACCATCATGG 60

|||||

QY 102 attcaagcaccgcacacagtcgggtgttctgtattctctccagaaatcactgctcag 161

|||||

Db 61 ATTCAGCACCGCACACAGTCCGGTGTTCCTGTTATTCCTCCAGAAATCACTGCTTCAG 120

|||||

QY 162 aatagtgaccacagaaactttcagccacgaaccttttcaactcaaacccctcgcaaaaat 221

|||||

Db 121 AATATGAGTCCACAGAACTTTTCAGCCACGACCTTTTCACTCAAGGCCCTTGCAGAAAAT 180

|||||

QY 222 tatttgcagaaaaatgaaaatcttagggactatcccgatcctgttttggaaattatgacct 281

|||||

Db 181 TATTTCGTAGAAAATGAAATCTTAGG----- 208

|||||

QY 282 ttcttttggagtattcttctcttttctcattgtttaaacacatatccaaggtttcccttta 341

|||||

```
Db 209 ----- 208
QY 342 tattcttcaggatataccattctgggctctgtttgttcattattctgagcctcc 401
Db 209 -----GGATATCCATTCCTGGGCTCTGTTTGTTCATTAATTCGGAGCCTCC 257
QY 402 taatgcagtgaaagaaacacacaaactctgataattgagcgaataatgaatt 461
Db 258 TAATTCGAGTGNAAAGAAAACACAAACTCTGTATATATATGAGCCGAATAATGAATT 317
QY 462 ttcttagtgccctgggagcaatagctgggaatcattctctcattctggtttccatcc 521
Db 318 TTCTTAGTGCCCTGGGAGCAATAGCTGGAAATCATCTCTCATATTTGGTTTCATCTAG 377
QY 522 atcaaaactacatttggtgttattctcaccaaaatagt 559
Db 378 ATCAAACTACTATTGTGGTTATCTCACCAAAATAGT 415

RESULT 8
BE107659/c
LOCUS
DEFINITION ot38b02.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1619019
ACCESSION AI002083
VERSION AI002083
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 484)
AUTHORS Greg Lennon, Ph.D.
TITLES Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Unpublished (1997)
MEDLINE Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgaps-@email.nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
```

```
6:791-806, 1996)
TAG_LIB=UI-R-BT1
TAG_TISSUE=testis
TAG_SEQ=ACGCAG*
BASE COUNT 165 a 96 c 96 g 125 t
ORIGIN
Query Match 38.8%; Score 295.2; DB 9; Length 484;
Best Local Similarity 76.7%; Pred. No. 2.7e-32;
Matches 376; Conservative 0; Mismatches 108; Indels 6; Gaps 1;
QY 258 agatcctgtttgggaattatgaccttttctttggaggttatctctcttcttctactttgtaa 317
Db 484 AGATCCTCTTTGGGAATCATGAACCTTCTCATTTGGAGTTGTTCCCTTTACCTTGGTAA 425
QY 318 aaccatatcaaggtttccctttattcttctcaggtatccattctgggctctgttt 377
Db 424 ACCCATACCCAAAGGTTCCCTTTATATTTCTCTAGGATATCTTCTGGGCTCTGCTT 365
QY 378 tttcattatttgagcctctcttaattgagtgaaagaaacacacagaactctga 437
Db 364 TGTTTCATTAACCTCTGGGACCTTTCTGATTGCTCTGAAAAAGAAAACTACAGACACTCTGA 305
QY 438 taatattgagccgaataatgaattttcttagtgccttgaggaatagctggaatcattc 497
Db 304 TAAAAATGAGCCAGCGCATGAATTTACTTAGTCCCTGGGAGCAGCAGCTGGATCATTC 245
QY 498 tcttcatttggtttcattcattctctgatacaaaactacatttggtgttattctcaccacaata 557
Db 244 TCCTCATATGTGCTTCTCTCTAGATGGAGAAATTCATCTGTGGCTATTCTCCAGATGGTA 185
QY 558 gtcagtgtgaagctgttactgtctcttcttctggaattttgattacattgacttcca 617
Db 184 TTCAGTGTGGTGCTATTACCACTCTATTCTTGGGATTTGATTATGTTGATGATCTCA 125
QY 618 gcattattgaaattattctctctctctctctctctctctctctctctctctctctctctct 677
Db 124 CGTTCGCTGAACACTCTTCATTCCTCTCTTTTCCTCGATTTTGGGGTCTACTCAGAGAAA 65
QY 678 gtgattgtgaacaattgttctgactgacactgagaaataaagatgtgttaaaactccta 737
Db 64 GTGGGGG-----ATGCTGTTGAATAGCAGCTGTACAAAGTAAACTGCAAAATTGAAAAA 11
QY 738 aaaaaaaa 747
Db 10 AAAAAAAA 1
RESULT 9
AI002083/c
LOCUS
DEFINITION ot38b02.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1619019
ACCESSION AI002083
VERSION AI002083.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 482)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLES National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Unpublished (1997)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgaps-@email.nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
```

found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/brp/image/image.html
Insert Length: 598 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 393.

FEATURES

Location/Qualifiers
1. .482
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1619019"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGAGCGGCCCAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."
157 a 85 c 99 g 141 t
BASE COUNT
ORIGIN
source
Query Match 35.7%; Score 271; DB 9; Length 482;
Best Local Similarity 71.8%; Pred. No. 6.3e-29;
Matches 479; Conservative 0; Mismatches 0; Indels 188; Gaps 2;
QY 66 tcaatcctttcaaatatcacgcacacatcatggttcacgacacgacacagtcg 125
DB 482 TCATCTCCCTTCAATATACCGACCATCATGATTCAGCACCGCACAGTCGG 423
QY 126 tgttctggtattctccagaaatcaatgcttcagaatatgagtcacagaaattcag 185
DB 422 TGTTCTGTTATTCCTCCAGAAATCACTGCTTCAGATATGATGATCCAGAACTTCAG 363
QY 186 ccagacatttcaactcaaaccttgcaaaaattattgctagaaaaatgaaatc 245
DB 362 CCAGGACCTTTTCACTCAAGCCCTTCGCAAAAATTTATTTGCTAGAAAAATGAATCT 303
QY 246 taggactatccagatcctgtttggaattagaccttttcttttggagttatcttctt 305
DB 302 TAGS----- 297
QY 306 tcaattgttaaaaccatatccaaagtttccctttatatttttttcaggatattcattct 365
DB 298 -----GGATATCATTTCT 286
QY 366 ggggctctgtttgttcatttaattctggagccttctctaattgagtgaaaaaacc 425
DB 285 GGGGCTCTGTTTGTTCATTAATCTGAGCCCTTCTTAATTCAGTGAAGAAAAACCA 226
QY 426 cagaaactctgataataattgagccgaaataatgaattttcttagtccctgggagcaatag 485
DB 225 CAGAAACTCT----- 214
QY 486 ctggaatcattctcctcaatttggtttcattcctagatcaaaactacatttggtatt 545
DB 215 -----GATCAAACTACATTTGTGTTATT 191
QY 546 ctcacaaaatagtcagtgtaagcgtgttactgtctctgttcttgggaattttgattacat 605
DB 190 CTCACCAAAATAGTCAGTGAAGCGTGTACTGTCTGTTCTTGGGAATTTGATTACAT 131
QY 606 tgaacttctgaagattgaattattcattctctgccttctcatttttgggggtgcc 665
DB 130 TGATGACTTTTTCAGCATTTATTGAATTAATTCATTTCTGCTTTTCTCAATTTTGGGGTGCC 71
QY 666 actcagaggattgattgaaatgtttgtagctagcactgtgagaaatgaagatgag 725

Db 70 ACTCAGAGGATTGTGATTGTGAACAATGTTGTGACTAGCACTGTGAGAAATAAGATGTG 11
QY 726 ttaaaat 732
DB 10 TTAATAAT 4
RESULT 10
AA411806/c 410 bp mRNA linear EST 12-AUG-1997
LOCUS z167a03.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:727372
DEFINITION 3' similar to SW:CD20_HUMAN P11836 B-LYMPHOCTE ANTIGEN CD20 ;,
mRNA sequence.
ACCESSION AA411806
VERSION AA411806.1 GI:2070377
KEYWORDS EST
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 410)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B.,
Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,
,T., Waterston,R. and Wilson,R.
WashU-Merck EST Project 1997
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 608 Std Error: 0.00
Seq primer: -4ml3 fwd. ET from Amersham
High quality sequence stop: 401.
Location/Qualifiers
1. .410
/organism="Homo sapiens"
/db_xref="GDB:5924284"
/db_xref="taxon:9606"
/clone="IMAGE:727372"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGAGCGGCCCAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."

FEATURES

source
BASE COUNT 147 a 68 c 76 g 119 t
ORIGIN
Query Match 35.2%; Score 267.4; DB 9; Length 410;
Best Local Similarity 96.1%; Pred. No. 2.2e-28;
Matches 274; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 178 acttcagccacgaccttttcaactcaagcccttcgcaaaattatttgcgtagaaaaat 237
DB 410 ACTTTCAGCCAGCAGCTTTTCAACTCAAGCCCTTTCGCAAAAATTTATTTGCTAGAAAAAT 351
QY 238 gaaatctttaggactatccagatcctctgtttggaattatgaactttcttttggagttat 297
DB 350 GAAAATCTTAGGGACTATCCAGATCCTGTTTGGAAATATGACCTTTTCTTTTGGAGTTAT 291

```

QY 298 ctctcttttcaacttggttaaaacacatacccaaggtttccctcttatattttcttcaggata 357
|||||
Db 290 CTTCTCTTTTTCACCTTGTTAAAAACCATATCCCAAGGTTTCCCTTTATATTTCTTTTCAGGATA 231
|||||

QY 358 tccattctggggctgtgtttgttcatttaattctggagcctctccttaattgcagtgaaga 417
|||||
Db 230 TCCATTCTGGGGCTGTGTTTGTTCATTAAATCTGGAGCCTTCCTAATTCAGTGAAGAAG 171
|||||

QY 418 aaaaacacagaaactctgataataattgagccgaataatgaatt 462
|||||
Db 170 AAAAACACAGAACTCTGGGAACTTTTCATTACATTGATGACTTT 126
|||||

RESULT 11
AA781801/c 389 bp mRNA linear EST 31-DEC-1998
LOCUS ai58g01.s1 Soares_testis_NHT Homo sapiens cDNA clone 1375056 3,
DEFINITION similar to SW:CD20_HUMAN FL1836 B-LYMPHOCYTE ANTIGEN CD20 ;, mRNA
sequence.
ACCESSION AA781801
VERSION AA781801.1 GI:2841132
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 389)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 983 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 311.
FEATURES
source
1..389
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="1375056"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - oligo(dT) primer [5',
TGTTCACCAATCTGAAGTGGAGCGCGCCAAATTTTGTGTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 140 a 66 c 68 g 115 t
ORIGIN

Query Match 32.3%; Score 245.4; DB 9; Length 389;
Best Local Similarity 95.8%; Pred. No. 2.6e-25;
Matches 252; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 200 actcaagcccttgcaaaaaattttctagaaaaatgaaatcttagggactatccag 259
|||||
Db 389 ACTCAAGCCCTTGCAAAAATTATTTCGTAGAAAATGAAATCTTAGGACTATCCAG 330
|||||

```

```

QY 260 atcctgttggaattatgacctttttcttttgagtgatcttccttttcacatttgtaaaa 319
|||||
Db 329 ATCCTGTTTGGAAATTATGACCTTTTCTTTTGGAGTTATCTTCTCCCTTTTACACCTTCTTAAAA 270
|||||

QY 320 ccataatccaaaggttccctcttatattttcttcaggatataccattctggggctctgttttg 379
|||||
Db 269 CCATATCCCAAGGTTTCCCTTTATATTTCTTTCAGGATATCCATTCTGGGGCTCTGTTTTG 210
|||||

QY 380 ttcattaaattctggagcctctccttaattgcagtgaaagaaaaaacacagaaactctgata 439
|||||
Db 209 TTCATTAAATCTCGAGCCTTCCTAATTCAGTGAAGAAACAAACACACAGAACTCTCGGA 150
|||||

QY 440 atattgagccgaataatgaatt 462
|||||
Db 149 ATTTTGATTACATTGATGACTTT 127
|||||

RESULT 12
AQ108532 538 bp DNA linear GSS 29-AUG-1998
LOCUS CIT-HSP-2379G7.TF CIT-HSP Homo sapiens genomic clone 2379G7, DNA
DEFINITION sequence.
ACCESSION AQ108532
VERSION AQ108532
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 538)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
Venter,J.C.
TITLE Use of a random human BAC End Sequence Database for Sequence-Ready
Map Building
JOURNAL Unpublished (1998)
COMMENT Other_GSSs: CIT-HSP-2379G7.TR
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.
FEATURES
source
1..538
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2379G7"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelobAC11; Site_1: HindIII; Site_2:
HindIII"
BASE COUNT 140 a 108 c 115 g 175 t
ORIGIN

Query Match 19.5%; Score 148.4; DB 12; Length 538;
Best Local Similarity 93.4%; Pred. No. 6.8e-12;
Matches 155; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 571 tttactgtctctgttcttggaattttgtattacattgatgcatttcagcattattgaatt 630
|||||
Db 152 TCTTCTTCTTCTTATACAGGAAATTTTGATTATGATGATTCAGCATTTATTGAATT 211
|||||

QY 631 attcattctctcctctcctcaattttgggtggccactcagagattgtgattgaaca 690
|||||

```


Db 169 AAAAAAGAGCGGCAATGAATTTACTAGTGTCTGGGAGCAGCAGCTGGAATCATTTCT 110

Qy 499 cctcacattgttccatccatagataaatacatttggttattctcaccacaaatag 558

Db 109 CCTCATATTTGGCTCCCTCTAGATGGGAATTCATCTGTGGCTATTCACAGATTGGTAT 50

Qy 559 tcagtgaaggctgttactgtctctgttcttctgggaattttgattacattg 607

Db 49 GCAGTGTGTGTATACCATTCATTCAATGGGATTTGGTTATGTTG 1

RESULT 15

BE638317/c

LOCUS

DEFINITION BE638317 601 bp mRNA linear EST 28-AUG-2000

musculus cDNA clone MTA03 5', mRNA sequence.

ACCESSION

VERSION BE638317

KEYWORDS BE638317.1 GI:9932060

SOURCE

ORGANISM

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 601)

Jiang, H., Li, L. and Lu, G.

Expressed sequence tags from mouse testis apoptosis subtraction

CDNA library

Unpublished (2000)

Contact: Jiang H

Department of Pharmacology

Hunan Medical University

88 Xiangya road, Changsha, Hunan 410078, China

Tel: 086-0731-4805036

Fax: 086-0731-4497661

Email: lj12@public.cs.hn.cn

Seq Primer: 17 Promoter Primer

High quality sequence stop: 601.

Location/Qualifiers

1..601

/organism="Mus musculus"

/strain="C57BL"

/db_xref="taxon:10090"

/clone="MTA03"

/clone.lib="mouse testis apoptosis subtraction cDNA

library"

/sex="male"

/tissue_type="testis"

/cell_type="Spermatogenesis cells"

/dev_stage="adult"

/lab_host="JM109"

/note="vector: pUCm-T"

BASE COUNT 174 a 158 c 146 g 123 t

ORIGIN

Query Match 17.8%; Score 135; DB 10; Length 601;
Best Local Similarity 74.0%; Pred. No. 4.7e-10;
Matches 171; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

Qy 454 aatgaattttcttagtgcctggagcaatagctgggaatctctcctcacatttggttt 513

Db 247 AATGAATTTACTAGTGTCTGGGAGCAGCAGCTGGAATCATTCCTCATATTGGCTT 188

Qy 514 catctagatcaaacatttattgttattctcaccacaaatagtcagtgtaaggctgt 573

Db 187 CCTCTAGATGGGAATTCATCTGGGCTATTCACAGATGGTATGCAGTGGGTAT 128

Qy 574 tactgtcctgttcttgggaattttgattacattgatgactttcagcattattgaattt 633

Db 127 TACCATTCTATTCAATGGGATTTGGTTATTTGATGATCTTCAGCATTGTTGAATCTT 68

Qy 634 catttctgccttctcattttgggtgpcactcagagattgtgattg 684

Search completed: August 30, 2002, 02:08:00
Job time: 5388 sec

Db 67 CATTTCCCTCTTTTTCCTCAGTTTTTGGGATGCTACACAGAGGAGGGGATG 17

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 27, 2002, 11:02:16 ; Search time 26.58 Seconds
(without alignments)
1301.692 Million cell updates/sec

Title: US-09-821-821-2

Perfect score: 1027

Sequence: 1 MDSSSTAHSVFLVFPPEITA.....SLPFSILGCHSDCDCEQCC 200

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phase:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1027	100.0	200	4	Q9H3V2
2	1021	99.4	200	4	Q9BZH1
3	202.5	19.7	205	4	Q9H3V3
4	202.5	19.7	220	4	Q9HY18
5	202.5	19.7	239	4	Q96JQ5
6	200.5	19.5	197	4	Q9P1S3
7	188.5	18.4	217	11	Q9D7Z9
8	187.5	18.3	217	11	Q9N08
9	185.5	18.1	214	4	Q96HJ5
10	183.5	17.9	225	4	Q96PG6
11	183.5	17.9	248	4	Q9H2W1
12	182.5	17.8	225	4	Q9HC76
13	180	17.5	226	11	Q9ES61
14	178.5	17.4	225	11	Q9N05
15	176	17.1	268	11	Q9N10
16	174	16.9	226	11	Q9D3F6

17	172.5	16.8	220	11	Q9D8E5	Q9d8e5 mus musculu
18	172.5	16.8	267	4	Q9NXJ0	Q9nxj0 homo sapien
19	171	16.7	178	4	Q9H3V1	Q9h3v1 homo sapien
20	171	16.7	679	4	Q96JA4	Q96ja4 homo sapien
21	168	16.4	249	4	Q9H2N3	Q9h2n3 homo sapien
22	167	16.3	268	11	Q9EQZ0	Q9eqz0 mus musculu
23	165	16.1	135	6	Q95LT4	Q95lt4 macaca fasc
24	160.5	15.6	213	11	Q920C4	Q920c4 mus musculu
25	160.5	15.6	240	4	Q9GZW8	Q9gzw8 homo sapien
26	158.5	15.4	247	11	Q9N07	Q9n07 mus musculu
27	156	15.2	244	11	Q9N09	Q9n09 mus musculu
28	155.5	15.1	247	11	Q9D0X1	Q9d0x1 mus musculu
29	154	15.0	244	11	Q9D8W9	Q9d8w9 mus musculu
30	148	14.4	249	4	Q9H2L1	Q9h2l1 homo sapien
31	145.5	14.2	176	11	Q9DC76	Q9dc76 mus musculu
32	145.5	14.2	250	4	Q9BY19	Q9by19 homo sapien
33	139.5	13.6	190	11	Q9EQV7	Q9eqv7 mus musculu
34	137.5	13.4	124	11	Q99MX6	Q99mx6 mus musculu
35	137.5	13.4	167	4	Q96PG7	Q96pg7 homo sapien
36	137.5	13.4	230	11	Q9EQY9	Q9eqy9 mus musculu
37	136	13.2	234	11	Q99N04	Q99n04 mus musculu
38	123	12.0	125	11	Q91ZY5	Q91zy5 mus musculu
39	108.5	10.6	215	11	Q9D2W6	Q9d2w6 mus musculu
40	106	10.3	307	16	Q9CES9	Q9ces9 lactococcus
41	103	10.0	175	11	Q9D7R4	Q9d7r4 mus musculu
42	103	10.0	455	5	Q19442	Q19442 caenorhabdi
43	100	9.7	152	4	Q14298	Q14298 homo sapien
44	96	9.3	387	16	Q9PNQ1	Q9pnq1 campylobact
45	95.5	9.3	355	17	Q96ZJ3	Q96zj3 sulfolobus

ALIGNMENTS

RESULT 1

Q9H3V2 ID Q9H3V2 PRELIMINARY; PRT; 200 AA.

AC Q9H3V2;

DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE MS4A5 (MS4A5 PROTEIN).

GN MS4A5.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=TESTIS;

RA Ishibashi K., Sasaki S., Marumo F.;

RT "Cloning of three CD20 homolog from human, putative calcium

channels";

RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=21295030; PubMed=11401424;

RA Liang Y., Tedder T.F.;

RT "Identification of a CD20-, FcepsilonRIbeta-, and HTM4-Related Gene

Family: Sixteen New MS4A Family Members Expressed in Human and

Mouse.";

RL Genomics 72:119-127(2001).

DR EMBL; AB013103; BAB18739.1; -

DR EMBL; AF237907; AAK37416.1; -

SQ SEQUENCE 200 AA; 22283 MW; F9282E5D15BC5514 CRC64;

Query Match 100.0%; Score 1027; DB 4; Length 200;

Best Local Similarity 100.0%; Pred. No. 1.2e-84;

Matches 200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDSSSTAHSVFLVFPPEITA...TSQSPLOKLFARKKILGTTQILFGIM 60

|||||

```
Db 1 MDSSTAHPVFLVFPPEITASEYESTELSATFTSQSPLOKLFARKMKILGTIQLFGIM 60
QY 61 TFSFGVIFLTLKPYPRPFIFLSGYPFWGSVLFNSGAFIAVKRKTETLILSRIM 120
Db 61 TFSFGVIFLTLKPYPRPFIFLSGYPFWGSVLFNSGAFIAVKRKTETLILSRIM 120
QY 121 NFLSALGAIAIGILLTFTGFILODNYICGYSHQNSQCKAVTVLFGILITLMTFSIIELFI 180
Db 121 NFLSALGAIAIGILLTFTGFILODNYICGYSHQNSQCKAVTVLFGILITLMTFSIIELFI 180
QY 181 SLPFSILGCHSEDCEQCC 200
Db 181 SLPFSILGCHSEDCEQCC 200

RESULT 2
QYBZH1 PRELIMINARY; PRT; 200 AA.
AC Q9BZH1;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE TESTIS-EXPRESSED TRANSMEMBRANE-4 PROTEIN.
GN TETM4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21092614; PubMed=111162526;
RA Hulet M.D., Pagler E., Hornby J.R., Hogarth P.M., Eyre H.J.,
BAKER E., Crawford J., Sutherland G.R., Ohms S.J., Parish C.R.;
RT "Isolation, tissue distribution, and chromosomal localization of a
RT novel testis-specific human four-transmembrane gene related to CD20
RL Blochem. Biophys. Res. Commun. 280:374-379(2001).
DR EMBL; AF321127; AAK01641.1; -.
KW Transmembrane.
SQ SEQUENCE 200 AA; 22249 MW; 7B282E5D15B25BF4 CRC64;

Query Match 99.4%; Score 1021; DB 4; Length 200;
Best Local Similarity 99.5%; Pred. No. 4.3e-84;
Matches 199; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDSSTAHPVFLVFPPEITASEYESTELSATFTSQSPLOKLFARKMKILGTIQLFGIM 60
Db 1 MDSSTAHPVFLVFPPEITASEYESTELSATFTSQSPLOKLFARKMKILGTIQLFGIM 60
QY 61 TFSFGVIFLTLKPYPRPFIFLSGYPFWGSVLFNSGAFIAVKRKTETLILSRIM 120
Db 61 TFSFGVIFLTLKPYPRPFIFLSGYPFWGSVLFNSGAFIAVKRKTETLILSRIM 120
QY 121 NFLSALGAIAIGILLTFTGFILODNYICGYSHQNSQCKAVTVLFGILITLMTFSIIELFI 180
Db 121 NFLSALGAIAIGILLTFTGFILODNYICGYSHQNSQCKAVTVLFGILITLMTFSIIELFI 180
QY 181 SLPFSILGCHSEDCEQCC 200
Db 181 SLPFSILGCHSEDCEQCC 200

RESULT 3
QYB3V3 PRELIMINARY; PRT; 205 AA.
AC Q9B3V3;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE MS4A4.
GN MS4A4.
OS Homo sapiens (Human).
```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21142397; PubMed=11245982;
RA Ishibashi K., Suzuki M., Sasaki S., Imai M.;
RT "Identification of a new multigene four-transmembrane family (MS4A)
RT related to CD20, Htm4 and beta subunit of the high-affinity Ige
RT receptor.";
RL Gene 264:87-93(2001).
DR EMBL; AB013102; BAB18738.1; -.
SQ SEQUENCE 205 AA; 21790 MW; FDB12899FCC551B3 CRC64;
```

```
Query Match 19.7%; Score 202.5; DB 4; Length 205;
Best Local Similarity 31.8%; Pred. No. 1.1e-10;
Matches 49; Conservative 31; Mismatches 71; Indels 3; Gaps 2;
```

```
QY 39 LQKLFAR-KMKILGTIQLFGIMTFSFGVIFLTLKPYPRPFIFLSGYPFWGSVLFIN 97
Db 20 LQKFLKGPVKLVGVVQILTALMSLSMGITMCMASNTYGSNPISVIGYITWGSVMFII 79
QY 98 SGAFIAVKRKTETLILSRIMNFLSALGAIAIGILLTF--GFILDQNYICGYSHQNSQ 155
Db 80 SGSLIAAGIRTKGLVRGSLGNMTSSVLAASGILINTFSLAFYSFHHPCNYCYGNSNN 139
QY 156 CKAVTVLFGLITLMTFSIIELFSLPFSILGC 189
Db 140 CHGTMSILMGDLGMVLLLSVLEFCIAVSLSAFGC 173
```

```
RESULT 4
QYB18 PRELIMINARY; PRT; 220 AA.
AC Q9B18;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MS4A4A PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=PLACENTA;
RX MEDLINE=21295030; PubMed=11401424;
RA Liang Y., Tedder T.F.;
RT "Identification of a cd20-, fcepsilonribeta-, and htm4-related gene
RT family: sixteen new ms4a family members expressed in human and
RT mouse.";
RL Genomics 72:119-127(2001).
DR EMBL; AF237912; AAK37594.1; -.
SQ SEQUENCE 220 AA; 23220 MW; 7648C762197C1179 CRC64;
```

```
Query Match 19.7%; Score 202.5; DB 4; Length 220;
Best Local Similarity 31.8%; Pred. No. 1.2e-10;
Matches 49; Conservative 31; Mismatches 71; Indels 3; Gaps 2;
```

```
QY 39 LQKLFAR-KMKILGTIQLFGIMTFSFGVIFLTLKPYPRPFIFLSGYPFWGSVLFIN 97
Db 35 LQKFLKGPVKLVGVVQILTALMSLSMGITMCMASNTYGSNPISVIGYITWGSVMFII 94
QY 98 SGAFIAVKRKTETLILSRIMNFLSALGAIAIGILLTF--GFILDQNYICGYSHQNSQ 155
Db 95 SGSLIAAGIRTKGLVRGSLGNMTSSVLAASGILINTFSLAFYSFHHPCNYCYGNSNN 154
QY 156 CKAVTVLFGLITLMTFSIIELFSLPFSILGC 189
Db 155 CHGTMSILMGDLGMVLLLSVLEFCIAVSLSAFGC 188
```



```
ID Q99N08 PRELIMINARY; PRT; 217 AA.
AC Q99N08;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MS4A6C PROTEIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-TOTAL FETUS;
RX MEDLINE=21295030; PubMed=11401424;
RA Liang Y., Tedder T.F.;
RT "Identification of a cd20-, fcepsilonribeta-, and htm4-related gene
family: sixteen new ms4a family members expressed in human and
mouse."
RL Genomics 72:119-127(2001).
DR EMBL: AF237910; AAK37419.1; -.
SQ SEQUENCE 217 AA; 23622 MW; 8BF521AF22DBB7BD CRC64;

Query Match 18.3%; Score 187.5; DB 11; Length 217;
Best Local Similarity 26.0%; Pred. No. 2.5e-09;
Matches 54; Conservative 41; Mismatches 80; Indels 33; Gaps 5;

QY 2 DSSPAHSPVFLVFPPEITASEYESTELSATFTSQPLQKLFARKMKILGTIQLFGIMT 61
   : : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 9 ETITSPNGINFPQK-----DESQPTQQRQDSLKKLHAEKIVIAIQCMAVTV 59

QY 62 FSGGVIFLTLKPYRPFIF----LSGYFPGWSVLFNSGAFLIAVKKTKTETILII 117
   : : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 60 LALGII--LASVPPVPVFNFSVLLKSGYPFGIFAGLFFIASGILSIITKRSKPLVDAS 117

QY 118 RIMNFLSALGAIAIILLTGF-----ILDQNYICG-YSHQNSOCKAV 159
   : : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 118 LTNILSVSFAFGVGIILISVSLAGLHPASEQKQSKLSLIEHDYIQPFYNSRSECAVT 177

QY 160 TVFLGILITMTFSIIEFLSIFLSIL 187
   : : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 178 KSLITGALSVMLIISVLELGLALLSAML 205

RESULT 9
Q96HJ5 PRELIMINARY; PRT; 214 AA.
AC Q96HJ5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE UNKNOWN (PROTEIN FOR MGC:14809).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-LUNG CARCINOMA;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL: BC008487; AAH08487.1; -.
SQ SEQUENCE 214 AA; 22933 MW; D30C276213DB8F3D CRC64;

Query Match 18.1%; Score 185.5; DB 4; Length 214;
Best Local Similarity 28.7%; Pred. No. 3.8e-09;
Matches 60; Conservative 39; Mismatches 85; Indels 25; Gaps 8;

QY 1 MDSSTAHSPVFLVFPPEITASEYESTELSATFTSQPLQKLFARKMKILGTIQLFGI 59
   : : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 11 LGSASAHGTPGSEAGPEELNTSVYQPDGS-----PDYQK---AKLQVLGAIIILNAA 60
```

```
QY 60 MTFSGVIFLFTLLKPY-----PRFPFIFLSGYFPGWSVLFNSGAFLIAVKKTKTETILII 115
   : : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 61 MILALGV-FLGSLQYPYHFKHFFFTFTVPIWGAFFCGSSGLSVVAGIKPRTWIQ 119

QY 116 LSRIMNFLSALGAIAIILLTGFILDQNYI--CCYSHQNSQ----CKAVTVFLGILIT 169
   : : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 120 NSFGMNIASATATLVGTAFSLNIAVNIQSLRSC---HSSSESDLCNMGSGISNGMVS 176

QY 170 LMTFSIIEFLSIFLSILGCHSEDCDCEQ 198
   : : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 177 LLLILLELCVITISTIAMWCNANCCNSRE 205

RESULT 10
Q96PG6 PRELIMINARY; PRT; 225 AA.
AC Q96PG6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MS4A6A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=21382044; PubMed=11486273;
RA Liang Y., Buckley T.R., Tu L., Langdon S.D., Tedder T.F.;
RT "Structural organization of the human MS4A gene cluster on Chromosome
11q12."
RL Immunogenetics 53:357-368(2001).
DR EMBL: AF354930; AAL07357.1; -.
SQ SEQUENCE 225 AA; 24331 MW; 12FBF5DCDC71B56A CRC64;

Query Match 17.9%; Score 183.5; DB 4; Length 225;
Best Local Similarity 26.4%; Pred. No. 6e-09;
Matches 55; Conservative 39; Mismatches 91; Indels 23; Gaps 6;

QY 1 MDSSTAHSPVFLVFPPEITASEYESTELSATFTSQPLQKLFARKMKILGTIQLFGIM 60
   : : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1 MTSQVPNETIIVLPSNVI--NFSQAEPETNOGQDSLKKLHAEIKVIGTIIQLCGMM 58

QY 61 TFSFGVIFLTLKP-YPRFPFIFL-SCYPFGWSVLFNSGAFLIAVKKTKTETILILSR 118
   : : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 59 VLSLGIILASASFSNFTQVSTLLNSAYPIGFFFIISGLSIATEKRLTKLLVHSSL 118

QY 119 IMNFLSALGAIAIILL-----TFGFILDQN-----YICGYSHQN---SOCKAV 159
   : : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 119 VGSILSALSALVGFIILSVKQATLNPASLQCELDKNIPTRSYVSYFVHDSLYTTDCVTA 178

QY 160 TVFLGILITMTFSIIEFLSIFLSIL 187
   : : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 179 KASLAGTILSLICTLLEFLCLAVLTAVL 206

RESULT 11
Q9H2W1 PRELIMINARY; PRT; 248 AA.
AC Q9H2W1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CDA01 (MS4A6A-POLYMORPH) (MS4A6A PROTEIN).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-PHEOCHROMOCYTOMA;
```



```
RA Li Y., Huang Q., Peng, Y, Song H., Yu Y., Xu S., Ren S., Chen Z.,
RA Han Z.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP .SEQUENCE FROM N.A.
RC TISSUE=LIVER, AND SPLEEN;
RX MEDLINE=21295030; PubMed=11401424;
RA Liang Y., Tedder T.F.;
RT "Identification of a CD20-, FcpsilonRIbeta-, and Htm4-Related Gene
RT Family; Sixteen New MS4A Family Members Expressed in Human and
RT Mouse.";
RL Genomics 72:119-127(2001).
DR EMBL; AF212240; AAG41780.1; -.
DR EMBL; AF286866; AAK37994.1; -.
DR EMBL; AF237908; AAK37417.1; -.
SQ SEQUENCE 248 AA; 26943 MW; 49892BC25D220A23 CRC64;

Query Match 17.9%; Score 183.5; DB 4; Length 248;
Best Local Similarity 26.4%; Pred. No. 6.6e-09;
Matches 55; Conservative 39; Mismatches 91; Indels 23; Gaps 6;

Qy 1 MDSSTAHPVLPVPPETASVESTLSATFTSQPQLKFAKMKILGTIQLFGIM 60
Db 1 MTSQVPNETIIVLPSNVI--NFSQAEKPEPTNQGDLSKKHLHAEIKVIGTIQLCGMM 58
Qy 61 TFSFGVIFLTLKLP-YPRPFPIFL-SGYFPWGSVLFNSGAFLIAVKRKTETLILSR 118
Db 59 VLSGLIILASFSFNFTQVTLNLSAYPIGPFPIISGSLSIATEKRLKLLVHSSL 118
Qy 119 IMNFSALGAIAGIILL-----TFGFILDQN-----YICGYSHQN---SQCKAV 159
Db 119 VGSILSALSALVGFIILSVKQATLNPSALQCELDKNNIPTRSYVSYFYHDSLYTDCYTA 178
Qy 160 TVLFGILITMTSIIELFISLPSIL 187
Db 179 KASLAGTSLMLICTLLEFCIAVLTAVL 206

RESULT 12
Q9HC76 PRELIMINARY; PRT; 225 AA.
AC Q9HC76;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE CD20-LIKE PRECURSOR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhang W., Wan T., He L., Yuan Z., Cao X.;
RT "Novel human CD20-like molecule.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF142409; AAG27920.1; -.
SQ SEQUENCE 225 AA; 24317 MW; 144D2EB7DC71B56A CRC64;

Query Match 17.8%; Score 182.5; DB 4; Length 225;
Best Local Similarity 26.4%; Pred. No. 7.4e-09;
Matches 55; Conservative 39; Mismatches 91; Indels 23; Gaps 6;

Qy 1 MDSSTAHPVLPVPPETASVESTLSATFTSQPQLKFAKMKILGTIQLFGIM 60
Db 1 MTSQVPNETIIVLPSNVI--NFSQAEKPEPTNQGDLSKKHLHAEIKVIGTIQLCGMM 58
Qy 61 TFSFGVIFLTLKLP-YPRPFPIFL-SGYFPWGSVLFNSGAFLIAVKRKTETLILSR 118
Db 59 VLSGLIILASFSFNFTQVTLNLSAYPIGPFPIISGSLSIATEKRLKLLVHSSL 118
Qy 119 IMNFSALGAIAGIILL-----TFGFILDQN-----YICGYSHQN---SQCKAV 159
```

```
Db 119 VGSILSALSALVGFIILSVKQATLNPSALQCELDKNNIPTRSYVSYFYHDSLYTDCYTA 178
Qy 160 TVLFGILITMTSIIELFISLPSIL 187
Db 179 KASLAGTSLMLICTLLEFCIAVLTAVL 206

RESULT 13
Q9ES61 PRELIMINARY; PRT; 226 AA.
AC Q9ES61;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CHANDRA PROTEIN (MS4A4B PROTEIN).
GN LY116.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SPLEEN;
RA Venkataraman C., Schaefer G., Schindler U.;
RT "Chandra, a novel four transmembrane protein differentially expressed
RT in helper type I lymphocytes.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=SPLEEN;
RX MEDLINE=21295030; PubMed=11401424;
RA Liang Y., Tedder T.F.;
RT "Identification of a CD20-, FcpsilonRIbeta-, and Htm4-Related Gene
RT Family; Sixteen New MS4A Family Members Expressed in Human and
RT Mouse.";
RL Genomics 72:119-127(2001).
DR EMBL; AF230640; AAG09739.1; -.
DR EMBL; AF237913; AAK37595.1; -.
DR MGD; MGI:1913083; LY116.
SQ SEQUENCE 226 AA; 24090 MW; 78C6ED3E830AA116 CRC64;

Query Match 17.5%; Score 180; DB 11; Length 226;
Best Local Similarity 26.1%; Pred. No. 1.2e-08;
Matches 43; Conservative 36; Mismatches 62; Indels 24; Gaps 4;

Qy 40 QKLFARKMKILGTIQLFGIMTFSFGVIFLTLKLPYPRPFPIISGYPFWGSVLFINS 99
Db 34 EKFLKGEPRVLGVQVIAINISLGIILITLFLSELPTSVMLKV---PIWGSIMFIVSG 90
Qy 100 AFLIAVKRKTETLII-----LSRIMNFSALGAIAGIILITFG--FILQNYICGYSHQN 153
Db 91 SLSIAAGVTPTKCLIVASLTNITSVLAATASIMGVSVAVGSQFPFRNY-----142
Qy 154 SQCKAVTVLFGILITMTSIIELFISLPSILSGHSEDCDCEQ 198
Db 143 -----TITKGLDVLMLIFNMLEFCIAVSVSAFGCEASCCNSRE 180

RESULT 14
Q99N05 PRELIMINARY; PRT; 225 AA.
AC Q99N05;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MS4A4D PROTEIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
```

Search completed: August 27, 2002, 11:05:06
Job time: 170 sec

```
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RX MEDLINE-211295030; PubMed=11401424;
RA Liang Y., Tedder T.F.;
RT "Identification of a cd20-, fcepsilonribeta-, and htm4-related gene
RT family: sixteen new ms4a family members expressed in human and
RT mouse.";
RL Genomics 72:119-127(2001).
DR EMBL; AF237915; AAK37598.1; -.
SQ SEQUENCE 225 AA; 23859 MW; 9D3B2678787039B4 CRC64;

Query Match 17.4%; Score 178.5; DB 11; Length 225;
Best Local Similarity 31.4%; Pred. No. 1.7e-08;
Matches 50; Conservative 35; Mismatches 61; Indels 13; Gaps 6;

QY 40 OKLFARKMKILGTQILGIMTFSEFGVIFLFTLLKPYPRPFIFLSGYPPWGSVLFINS 99
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
34 EKFLGGEKVLGAQVMIAFINFSLGIII--LNKVSERFMSVLLLA-PFWGSIIMFISG 90
QY 100 AFLIAVKRKTTETIIILSRIMNFLSALGAIAAGIILLTFGEFLDONICGYSHQNSOCKAV 159
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
91 SLSIAAGVKPTKAMITISLSVNTISSVLAAVASII---GVI---SVISGVFRQPRSQPAI 144
QY 160 TVLFGILITLMTFSIIELFISLPSILGCHSEDCDCEQ 198
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
145 A--SLDVLMTIL--NMLEFCIAVSVAFGCKACCCNSE 179

RESULT 15
Q99N10 PRELIMINARY; PRT; 268 AA.
AC Q99N10;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MS4A8B PROTEIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=COLON;
RX MEDLINE-21295030; PubMed=11401424;
RA Liang Y., Tedder T.F.;
RT "Identification of a cd20-, fcepsilonribeta-, and htm4-related gene
RT family: sixteen new ms4a family members expressed in human and
RT mouse.";
RL Genomics 72:119-127(2001).
DR EMBL; AF237906; AAK37415.1; -.
SQ SEQUENCE 268 AA; 28568 MW; 490C79BF3BA1579C CRC64;

Query Match 17.1%; Score 176; DB 11; Length 268;
Best Local Similarity 27.1%; Pred. No. 3.3e-08;
Matches 60; Conservative 36; Mismatches 91; Indels 34; Gaps 9;

QY 1 MDSSTHSP----VFLVFPPEI-----TASE---YESTELSATTF-----TQS 37
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
17 MNTVTSPPGMANSVYVAPPNSVYPVPGTVPQMPIYPSNQPVHVISGHLPGIVPAMTEP 76
QY 38 PLQKLFARKMKILGTQILGIMTFSEFGVIFLFTLLKPYPRPFIFLSGYPPWGSVLFIN 97
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
77 PAQRVL--KKGQVLGAQILGLVHIGLSIMITNLFSHYT--PVSLYGGFPFWGGIWFII 133
QY 98 SGAFLIAVKRKTTETIIILSRIMNFLSALGAIAAGIILLTFGEFLDONICGY---SHQN 153
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
134 SGSLVAAEFQPNPCLNGSVGLNIFSAICSAVGIML-----FITDISISSGIIYPSYP 189
QY 154 SQKAVTVLFLGILITLMTFSIIELFISLPSILGCHSEDC 194
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
190 YQENLGVRTGVAISSVLLIFCLLELSIASVSSHFGQVACC 230
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 27, 2002, 11:02:16 ; Search time 27.48 Seconds
(without alignments)
177.770 Million cell updates/sec

Title: US-09-821-821-2
Perfect score: 1027
Sequence: 1 MDSSTAHPVLVFPPEITA.....SLPFSILGCHSDCDCEQCC 200

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	202.5	19.7	239	2	US-08-916-902A-1
2	202.5	19.7	239	2	US-09-213-389-1
3	181.5	17.7	214	1	US-08-318-492-4
4	181.5	17.7	214	1	US-08-707-340-4
5	181.5	17.7	214	2	US-08-916-902A-3
6	181.5	17.7	214	2	US-08-994-578-4
7	181.5	17.7	214	2	US-09-213-389-3
8	158	15.4	243	1	US-07-869-933-29
9	158	15.4	243	1	US-07-869-933-33
10	158	15.4	243	1	US-08-201-879A-4
11	158	15.4	243	2	US-08-916-902A-4
12	158	15.4	243	2	US-09-213-389-4
13	158	15.4	243	4	US-09-103-663-29
14	158	15.4	243	4	US-09-103-663-33
15	158	15.4	246	1	US-07-869-933-23
16	158	15.4	246	4	US-09-103-663-23
17	157.5	15.3	235	1	US-07-869-933-34
18	157.5	15.3	235	1	US-08-201-879A-5
19	157.5	15.3	235	4	US-09-103-663-34
20	123.5	12.0	244	1	US-07-869-933-32
21	123.5	12.0	244	1	US-08-201-879A-3
22	123.5	12.0	244	4	US-09-103-663-32
23	87.5	8.5	956	2	US-08-897-443-3
24	82	8.0	327	4	US-08-748-506-24
25	80	7.8	371	2	US-08-928-692-20
26	79.5	7.7	241	3	US-08-808-148-1
27	79.5	7.7	241	4	US-09-020-956-114

ALIGNMENTS

RESULT 1

US-08-916-902A-1

; Sequence 1, Application US/08916902A

; Patent No. 5871930

; GENERAL INFORMATION:

; APPLICANT: Bandman, Olga

; APPLICANT: Lal, Preeti

; TITLE OF INVENTION: HIGH AFFINITY IMMUNOGLOBULIN E

; TITLE OF INVENTION: RECEPTOR-LIKE PROTEIN

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FASTSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/916,902A

; FILING DATE: Herewith

; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0371 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-855-0555

; TELEFAX: 415-845-4166

; TELEX:

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 239 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: BRAINOT04

; CLONE: 927955

US-08-916-902A-1

Sequence 114, App

Sequence 114, App

Sequence 126, App

Sequence 4, Appli

Sequence 1, Appli

Sequence 14, Appl

Sequence 14, Appl

Sequence 17, Appl

Sequence 13, Appl

Sequence 14, Appl

Sequence 2, Appli

Sequence 20, Appl


```

US-08-916-902A-3
; Sequence 3, Application US/08916902A
; Patent No. 5871930
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HIGH AFFINITY IMMUNOGLOBULIN E
; TITLE OF INVENTION: RECEPTOR-LIKE PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/916,902A
; FILING DATE: Herewith
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0371 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 214 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 561639
US-08-916-902A-3

```

Query Match	17.7%	Score	181.5;	DB	2;	Length	214;
Best Local Similarity	28.2%;	Pred.	No. 5.2e-13;				
Matches	59;	Conservative	38;	Mismatches	87;	Indels	25; Gaps
							8;
Qy	1	MDSSTAH-SPVFLVFPPEITASEYESTELSATTFSTQSPLOKLFAARKMKILGTIOILFCI	59	:	: :	:	:
Db	11	LGSASAHGTPSGETCPPELNTSVYH-----PINGSPTYQK---AKLOVLGAIGAILNNA	60	:	:	:	:
Qy	60	MTFSEGVIFLFTLLKPX----PRPFPIFLSGYPFWGSLVFI NSGAFLIAVRKRKTTEII I	115	:	:	:	:
Db	61	MILALGV-FLGSLQPYPHFQRHFPPFTVTGYPIWGAVFFCGSSG TLSVVAGIKPTRWTIQ	119	:	:	:	:
Qy	116	LSRINNFLSALUGAIGIILLFFGFILDQNYI--CCYSHQNSQ----KRAVTVLFGLILIT	169	:	:	:	:
Db	120	NSFGNNIASATIALTGTAFLSUNIAVNTQSURSC---HSSSEPDLCNMYGMSISNGMVS L	176	:	:	:	:
Qy	170	LMTFSIIIEFLSFPSILGSHESDCDCEQ	198	:	:	:	:
Db	177	LILTLLELCYTISTIAMWCNANCNSRE	205	:	:	:	:
RESULT	6						
US-08-994-578-4							
; Sequence 4, Application US/08994578							


```

; APPLICANT: KINET, Jean-Pierre
; TITLE OF INVENTION: ISOLATION, CHARACTERIZATION, AND USE OF
; TITLE OF INVENTION: THE HUMAN B SUBUNIT OF THE HIGH AFFINITY RECEPTOR FOR
; TITLE OF INVENTION: IMMUNOGLOBULIN
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/07/869,933
; FILING DATE: 19920416
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 40399/154 NIHD
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 243 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; STRAIN: beta subunit
; US-07-869-933-29

Query Match 15.4%; Score 158; DB 1; Length 243;
Best Local Similarity 28.0%; Pred. No. 3e-10;
Matches 45; Conservative 28; Mismatches 72; Indels 16; Gaps 3;

QY 36 QSPLOKLFARKMKILGTQILFGIMTFSGVIFLTLKP--YPRFPFIPLSGYPFWGAV 93
Db 47 QQTWQSFLLKKELEFGVTQVLGLICFCFVVCSTLQTSDFDDEVLVLLYRAGYPPFWGAV 106
QY 94 LFINSGAFLLIAYKRKTTETLILSRIMNFLSALGAIAIGIILLTFGLDQNYICGYSHQN 153
Db 107 LFLVLSGFLSIMSKNTLYLVKSLGANIVSSIAAGLGIAILLNLNLSNAYM----- 159
QY 154 SOCKAVT-----VLFGILITLMTFSIIEFLPSIL 187
Db 160 NYCKDITDDGCFVTSFTELVLMLLFTLILAFCSAVLLII 200

RESULT 9
US-07-869-933-33
; Sequence 33, Application US/07869933
; Patent No. 5770396
; GENERAL INFORMATION:
; APPLICANT: KINET, Jean-Pierre
; TITLE OF INVENTION: ISOLATION, CHARACTERIZATION, AND USE OF
; TITLE OF INVENTION: THE HUMAN B SUBUNIT OF THE HIGH AFFINITY RECEPTOR FOR
; TITLE OF INVENTION: IMMUNOGLOBULIN
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA

; APPLICANT: KINET, Jean-Pierre
; TITLE OF INVENTION: ISOLATION, CHARACTERIZATION, AND USE OF
; TITLE OF INVENTION: THE HUMAN B SUBUNIT OF THE HIGH AFFINITY RECEPTOR FOR
; TITLE OF INVENTION: IMMUNOGLOBULIN
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/07/869,933
; FILING DATE: 19920416
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 40399/154 NIHD
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 243 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; STRAIN: beta subunit
; US-07-869-933-29

Query Match 15.4%; Score 158; DB 1; Length 243;
Best Local Similarity 28.0%; Pred. No. 3e-10;
Matches 45; Conservative 28; Mismatches 72; Indels 16; Gaps 3;

QY 36 QSPLOKLFARKMKILGTQILFGIMTFSGVIFLTLKP--YPRFPFIPLSGYPFWGAV 93
Db 47 QQTWQSFLLKKELEFGVTQVLGLICFCFVVCSTLQTSDFDDEVLVLLYRAGYPPFWGAV 106
QY 94 LFINSGAFLLIAYKRKTTETLILSRIMNFLSALGAIAIGIILLTFGLDQNYICGYSHQN 153
Db 107 LFLVLSGFLSIMSKNTLYLVKSLGANIVSSIAAGLGIAILLNLNLSNAYM----- 159
QY 154 SOCKAVT-----VLFGILITLMTFSIIEFLPSIL 187
Db 160 NYCKDITDDGCFVTSFTELVLMLLFTLILAFCSAVLLII 200

RESULT 9
US-07-869-933-33
; Sequence 33, Application US/07869933
; Patent No. 5770396
; GENERAL INFORMATION:
; APPLICANT: KINET, Jean-Pierre
; TITLE OF INVENTION: ISOLATION, CHARACTERIZATION, AND USE OF
; TITLE OF INVENTION: THE HUMAN B SUBUNIT OF THE HIGH AFFINITY RECEPTOR FOR
; TITLE OF INVENTION: IMMUNOGLOBULIN
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
```

```

; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/07/869,933
; FILING DATE: 19920416
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 40399/154 NIHD
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 243 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: rat
; STRAIN: FCRI beta subunit
; US-07-869-933-33

Query Match 15.4%; Score 158; DB 1; Length 243;
Best Local Similarity 28.0%; Pred. No. 3e-10;
Matches 45; Conservative 28; Mismatches 72; Indels 16; Gaps 3;

QY 36 QSPLOKLFARKMKILGTQILFGIMTFSGVIFLTLKP--YPRFPFIPLSGYPFWGAV 93
Db 47 QQTWQSFLLKKELEFGVTQVLGLICFCFVVCSTLQTSDFDDEVLVLLYRAGYPPFWGAV 106
QY 94 LFINSGAFLLIAYKRKTTETLILSRIMNFLSALGAIAIGIILLTFGLDQNYICGYSHQN 153
Db 107 LFLVLSGFLSIMSKNTLYLVKSLGANIVSSIAAGLGIAILLNLNLSNAYM----- 159
QY 154 SOCKAVT-----VLFGILITLMTFSIIEFLPSIL 187
Db 160 NYCKDITDDGCFVTSFTELVLMLLFTLILAFCSAVLLII 200

RESULT 10
US-08-201-879A-4
; Sequence 4, Application US/08201879A
; Patent No. 5807988
; GENERAL INFORMATION:
; APPLICANT: KINET, Jean-Pierre
; APPLICANT: JOUVIN, Marie-Helene
; TITLE OF INVENTION: ISOLATION, CHARACTERIZATION, AND USE OF
; TITLE OF INVENTION: THE HUMAN B SUBUNIT OF THE HIGH AFFINITY RECEPTOR FOR
; TITLE OF INVENTION: IMMUNOGLOBULIN E
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/201,879A
; FILING DATE: 24-FEB-1994
```



```

; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 243 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 204117
US-09-213-389-4

Query Match          15.4%; Score 158; DB 2; Length 243;
Best Local Similarity 28.0%; Pred. No. 3e-10;
Matches 45; Conservative 28; Mismatches 72; Indels 16; Gaps 3;

QY 36 QSPLOKLFARKMKILGTIQLFGIMTFSGVIFLTLKP--YPRPFIFLSCYPFWSV 93
DB 47 QQTWSFLKKELEFLGVTVQLVGLICLCFGTVVCSLTQTSDFDEDEVLVLLYRAGYFPGAV 106
QY 94 LFINSAGFLIAYKRTTETLILSRIMNFLSALCAIAGIILLTFGLDQNYICGYSHQN 153
DB 107 LFLVSGFLSIMSRKNTLYLRGSLGANIVSSIAAGLGIAILLNLNLSNNSAYM----- 159
QY 154 SOCKAVT-----VLFGLILITLMTFSIIEFLISLPFSIL 187
DB 160 NYCKDITEDDGCFTVSFTITELVLMLEFLTLAFCSAVLLII 200

RESULT 13
US-09-103-663-29
; Sequence 29, Application US/09103663D
; Patent No. 6171803
; GENERAL INFORMATION:
; APPLICANT: Kinet et al.
; TITLE OF INVENTION: Isolation, characterization, and use of the human beta
; FILE OF INVENTION: subunit of the high affinity receptor for
; FILE REFERENCE: 50490
; CURRENT APPLICATION NUMBER: US/09/103,663D
; EARLIER FILING DATE: 1998-06-23
; EARLIER FILING DATE: 1992-04-16
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 29
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-103-663-29

Query Match          15.4%; Score 158; DB 4; Length 243;
Best Local Similarity 28.0%; Pred. No. 3e-10;
Matches 45; Conservative 28; Mismatches 72; Indels 16; Gaps 3;

QY 36 QSPLOKLFARKMKILGTIQLFGIMTFSGVIFLTLKP--YPRPFIFLSCYPFWSV 93
DB 47 QQTWSFLKKELEFLGVTVQLVGLICLCFGTVVCSLTQTSDFDEDEVLVLLYRAGYFPGAV 106
QY 94 LFINSAGFLIAYKRTTETLILSRIMNFLSALCAIAGIILLTFGLDQNYICGYSHQN 153
DB 107 LFLVSGFLSIMSRKNTLYLRGSLGANIVSSIAAGLGIAILLNLNLSNNSAYM----- 159
QY 154 SOCKAVT-----VLFGLILITLMTFSIIEFLISLPFSIL 187
DB 160 NYCKDITEDDGCFTVSFTITELVLMLEFLTLAFCSAVLLII 200

RESULT 14
US-09-103-663-33
; Sequence 33, Application US/09103663D
; Patent No. 6171803
; GENERAL INFORMATION:
; APPLICANT: Kinet et al.
; TITLE OF INVENTION: Isolation, characterization, and use of the human beta
; FILE OF INVENTION: subunit of the high affinity receptor for
; FILE REFERENCE: 50490
; CURRENT APPLICATION NUMBER: US/09/103,663D
; EARLIER FILING DATE: 1998-06-23
; EARLIER FILING DATE: 1992-04-16
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Rattus sp.
US-09-103-663-33

Query Match          15.4%; Score 158; DB 4; Length 243;
Best Local Similarity 28.0%; Pred. No. 3e-10;
Matches 45; Conservative 28; Mismatches 72; Indels 16; Gaps 3;

QY 36 QSPLOKLFARKMKILGTIQLFGIMTFSGVIFLTLKP--YPRPFIFLSCYPFWSV 93
DB 47 QQTWSFLKKELEFLGVTVQLVGLICLCFGTVVCSLTQTSDFDEDEVLVLLYRAGYFPGAV 106
QY 94 LFINSAGFLIAYKRTTETLILSRIMNFLSALCAIAGIILLTFGLDQNYICGYSHQN 153
DB 107 LFLVSGFLSIMSRKNTLYLRGSLGANIVSSIAAGLGIAILLNLNLSNNSAYM----- 159
QY 154 SOCKAVT-----VLFGLILITLMTFSIIEFLISLPFSIL 187
DB 160 NYCKDITEDDGCFTVSFTITELVLMLEFLTLAFCSAVLLII 200

RESULT 15
US-07-869-933-23
; Sequence 23, Application US/07869933
; Patent No. 5770396
; GENERAL INFORMATION:
; APPLICANT: KINET, Jean-Pierre
; TITLE OF INVENTION: ISOLATION, CHARACTERIZATION, AND USE OF
; TITLE OF INVENTION: THE HUMAN B SUBUNIT OF THE HIGH AFFINITY RECEPTOR FOR
; TITLE OF INVENTION: IMMUNOGLOBULIN
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/869,933
; FILING DATE: 19920416
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 40399/154 NIH
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)883-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246 amino acids

```

; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-869-933-23

Query Match 15.4%; Score 158; DB 1; Length 246;
Best Local Similarity 28.0%; Pred. No. 3e-10;
Matches 45; Conservative 28; Mismatches 72; Indels 16; Gaps 3;
QY 36 QSPLOKLPARKKILGTIQILFGIMTFSGVIFLFTLLKP--YPRPFIPLSGYPPWGSV 93
DB 50 QQTWQSFLLKKEFLGVTQVLVGLICLCFCGVVVCSTLQTSDFDDEVLVLLYRAGYPPWGA 109
QY 94 LFINSGAFLIAVRKKTETIILSRIMNPLSALGAIAGIILLTFGFIIDQNYICGYSHQN 153
DB 110 LEVLGFLSINSEKNTLYLVKSLGANIVSSIAAGLGIALLILNLSNAYM----- 162
QY 154 SOCKAVT-----VLFLGILITIMTFSEIPLPSIL 187
DB 163 NYCKDITEDDGCFTVSPITELVLMLLFLTLAFCSAVLLII 203

Search completed: August 27, 2002, 11:02:53
Job time: 37 sec

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M62541; AAA37394.1; -;
EMBL; AK017903; BAB30996.1; -;
PRL; A30558; A30558.
MGI:88321; Ms4a2.
B-cell; Transmembrane; Phosphorylation.
FT DOMAIN 1 44 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 45 65 POTENTIAL.
FT TRANSMEM 69 89 POTENTIAL.
FT TRANSMEM 112 132 POTENTIAL.
FT TRANSMEM 183 203 POTENTIAL.
FT DOMAIN 204 291 CYTOPLASMIC (POTENTIAL).
SEQUENCE 291 AA; 31958 MW; DP478ECD2C5C16FC CRC64;

Query Match 16.8%; Score 173; DB 1; Length 291;
Best Local Similarity 29.4%; Pred. No. 2.7e-07;
Matches 53; Conservative 29; Mismatches 64; Indels 34; Gaps 7;

QY 32 TFSQSPQLKLFARKMKILGTITQILFGIMTFSGVIFLFTLLKYPFRPFIFLS-GYFEW 90
DB 27 TSSLVGPQSFMRKSKALGAVQIMNGLFHITLGLGGL-----LMIPTGVFAPICLSVWYPLW 82
QY 91 GSVLFINSGLFIIVKRTTETLILSR-IMNFLSALGAIAGIIL-----LTFGILD 142
DB 83 GGIMYIISGSLAAAEKTSKSLVKAVKVINSSLSLFAAISGILSMDIILNLTSLHFLK 142
QY 143 QN-----YIC---GYSHONS-----OCKAVTVLFLGILITLMTFSIIEFLIS 181
DB 143 MRRLIQTSPKYVDIYDCEPSNSSEKNSPSTQYCNISQISVFLGILSAMLISAFQKLVT 202

RESULT 2
CD20_HUMAN STANDARD; PRT; 297 AA.
AC P11836; P08984; Q13963;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE B-lymphocyte antigen CD20 (B-lymphocyte surface antigen B1) (Leu-16) (Bp35).
GN MS4A1 OR CD20.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88124792; PubMed=2448768;
RA Stamenkovic I., Seed B.;
RT "Analysis of two cDNA clones encoding the B lymphocyte antigen CD20 (B1, Bp35), a type III integral membrane protein.";
RL J. Exp. Med. 167:1975-1980(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88124792; PubMed=2448768;
RA Tedder T.F., Streuli M., Schlossman S.F., Saito H.;
RT "Isolation and structure of a cDNA encoding the B1 (CD20) cell-surface antigen of human B lymphocytes";
PL Proc. Natl. Acad. Sci. U.S.A. 85:208-212(1988).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=89176281; PubMed=2466899;
RA Tedder T.F., Klejman G., Schlossman S.F., Saito H.;
RT "Structure of the gene encoding the human B lymphocyte differentiation antigen CD20 (B1).";
RL J. Immunol. 142:2560-2568(1989).
RN [4]
RP SEQUENCE FROM N.A.

RX MEDLINE=88283639; PubMed=2456210;
RA Einfeld D.A., Brown J.P., Valentine M.A., Clark E.A., Ledbetter J.A.;
RT "Molecular cloning of the human B cell CD20 receptor predicts a hydrophobic protein with multiple transmembrane domains.";
RL EMBO J. 7:711-717(1988).
CC -!- FUNCTION: THIS PROTEIN MAY BE INVOLVED IN THE REGULATION OF B-CELL ACTIVATION AND PROLIFERATION.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- PTM: PHOSPHORYLATED. MIGHT BE FUNCTIONALLY REGULATED BY PROTEIN KINASE(S).
CC -!- SIMILARITY: BELONGS TO THE MS4A FAMILY.
CC -!- DATABASE: NAME-PROW; NOTE-CD guide CD20 entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd20.htm".
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X12530; CAA31046.1; -;
EMBL; M27394; AAA35581.1; -;
EMBL; M27395; -; NOT_ANNOTATED_CDS.
EMBL; L23419; AAA88911.1; -;
EMBL; L23415; AAA88911.1; JOINED.
EMBL; L23416; AAA88911.1; JOINED.
EMBL; L23417; AAA88911.1; JOINED.
EMBL; L27203; CAA30179.1; -;
EMBL; X07204; CAA30180.1; -;
PIR; A27400; A27400.
PIR; J10042; J10042.
PIR; A30586; A30586.
PIR; S00387; S00387.
MIM; 112210; -;
KW B-cell; Transmembrane; Phosphorylation.
FT DOMAIN 1 63 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 64 84 POTENTIAL.
FT TRANSMEM 85 105 POTENTIAL.
FT TRANSMEM 121 141 POTENTIAL.
FT TRANSMEM 189 209 POTENTIAL.
FT DOMAIN 210 297 CYTOPLASMIC (POTENTIAL).
FT DISULFID 81 167 PROBABLE.
FT DISULFID 167 183 PROBABLE.
FT DISULFID 111 220 PROBABLE.
FT CONFLICT 13 13 P -> L (IN REF. 4).
FT CONFLICT 71 71 M -> I (IN REF. 3).
SQ SEQUENCE 297 AA; 33077 MW; AC5420F8B626BDD1 CRC64;

Query Match 16.5%; Score 169; DB 1; Length 297;
Best Local Similarity 30.3%; Pred. No. 5.9e-07;
Matches 54; Conservative 27; Mismatches 63; Indels 34; Gaps 7;
QY 34 STQSPQLKLFARKMKILGTITQILFGIMTFSGVIFLFTLLKYPFRPFIFLSGYPFWGS 92
DB 36 SLVGPQSFMRKSKALGAVQIMNGLFHITLGLGGL-----LMIPAGIYAPICVTVWYPLWGG 91
QY 93 VLFINSGLFIIVKRTTETLILSRIMNFLSALGAIAGIIL-----T 136
DB 92 IMYIISGSLAAAEKTSKSLVKAVKVINSSLSLFAAISGILSMDIILNLTSLHFLKMS 151
QY 137 FGFILDQN-YICG-----SHONS-----OCKAVTVLFLGILITLMTFSII-ELFIS 181
DB 152 LNFIRHTPYININCEPANPSEKNSPSTQYCNISQISVFLGILSAMLISAFQELVIA 209
RESULT 3
ID FCEB_RAT STANDARD; PRT; 243 AA.
AC P13386;
DT 01-JAN-1990 (Rel. 13, Created)


```

Db 117 YLVKSGUGANVSSIAAGTGIAMLIL-----NLTNFAIMNN-CRNVTDDGCFVAST 169
QY 165 GILITLMTFSIIELFIS 181
Db 170 TELVLMFLTILAFCS 186

RESULT 5
FCBE_HUMAN
ID FCBE_HUMAN STANDARD; PRT; 244 AA.
AC Q01362;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE High affinity immunoglobulin epsilon receptor beta-subunit (FCERI)
DE (Ige Fc receptor, beta-subunit) (Fc epsilon receptor I beta-chain).
GN MS4A2 OR FCER1B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Cranata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92316966; PubMed=1535625;
RA Kuester H., Zhang L., Brini A.T., Macglashan D.W., Kinet J.-P.;
RT "The gene and cDNA for the human high affinity immunoglobulin E
RT receptor beta chain and expression of the complete human receptor.";
RL J. Biol. Chem. 267:12782-12787(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92339505; PubMed=1386024;
RA Maekawa K., Imagawa N., Tanaka Y., Harada S.;
RT "Determination of the sequence coding for the beta subunit of the
RT human high-affinity Ige receptor.";
RL FEBS Lett. 302:161-165(1992).
RN [3]
RP VARIANT GLY-237.
RX MEDLINE=96414302; PubMed=8817330;
RA Hill M.R., Cookson W.O.;
RT "A new variant of the beta subunit of the high-affinity receptor for
RT Immunoglobulin E (Fc epsilon RI-beta E237G): associations with
RT measures of atopy and bronchial hyper-responsiveness.";
RL Hum. Mol. Genet. 5:959-962(1996).
RN [4]
RP VARIANT GLY-237.
RX MEDLINE=96440420; PubMed=8842731;
RA Shirakawa T., Mao X.-Q., Sasaki S., Enomoto T., Kawai M., Morimoto K.,
RA Hopkin J.;
RT "Association between atopic asthma and a coding variant of Fc-epsilon-
RT RI-beta in a Japanese population.";
RL Hum. Mol. Genet. 5:1129-1130(1996).
RN [5]
RP ERRATUM.
RX MEDLINE=97123518; PubMed=8968765;
RA Shirakawa T., Mao X.-Q., Sasaki S., Enomoto T., Kawai M., Morimoto K.,
RA Hopkin J.;
RL Hum. Mol. Genet. 5:2068-2068(1996).
CC [-] FUNCTION: BINDS TO THE FC REGION OF IMMUNOGLOBULINS EPSILON. HIGH
CC AFFINITY RECEPTOR. RESPONSIBLE FOR INITIATING THE ALLERGIC
CC RESPONSE. BINDING OF ALLERGEN TO RECEPTOR-BOUND IGE LEADS TO CELL
CC ACTIVATION AND THE RELEASE OF MEDIATORS (SUCH AS HISTAMINE)
CC RESPONSIBLE FOR THE MANIFESTATIONS OF ALLERGY. THE SAME RECEPTOR
CC ALSO INDUCES THE SECRETION OF IMPORTANT LYMPHOKINES.
CC [-] SUBUNIT: TETRAMER OF AN ALPHA CHAIN, A BETA CHAIN, AND TWO
CC DISULFIDE LINKED GAMMA CHAINS.
CC [-] SUBCELLULAR LOCATION: Integral membrane protein.
CC [-] TISSUE SPECIFICITY: FOUND ON THE SURFACE OF MAST CELLS AND
CC BASOPHILS.
CC [-] POLYMORPHISM: VARIANT GLU-237 HAS BEEN FOUND TO BE PRESENT IN ABOUT
CC 5.3% OF A 1004 INDIVIDUALS POPULATION SAMPLE IN AUSTRALIA. IT
CC SEEMS TO BE A RISK FACTOR FOR ATOPIC DERMATITIS AND ASTHMA.
CC [-] SIMILARITY: BELONGS TO THE M5A4 FAMILY.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D10583; BAA01440.1; -.
CC DR EMBL; M89796; AAA60269.1; -.
CC DR PIR; S21154; S21154.
CC DR PIR; A42806; A42806.
CC DR MIM; 147138; -.
CC KW Ige-binding protein; Receptor; Transmembrane; Polymorphism.
CC FT DOMAIN 1 59 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 60 79 POTENTIAL.
CC FT DOMAIN 80 97 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 98 117 POTENTIAL.
CC FT DOMAIN 118 130 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 131 150 POTENTIAL.
CC FT DOMAIN 151 180 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 181 200 POTENTIAL.
CC FT DOMAIN 201 244 CYTOPLASMIC (POTENTIAL).
CC FT VARIANT 237 237 E -> G.
CC FT SEQUENCE 244 AA; 26533 MW; CE523102D5F567AF CRC64;
CC SQ /FTID=VAR_003965.

Query Match 12.08; Score 123.5; DB 1; Length 244;
Best Local Similarity 25.6%; Pred. No. 0.0023;
Matches 50; Conservative 33; Mismatches 93; Indels 19; Gaps 7;

QY 9 PVFLVFPPEITASEVSELTSTFTSQPIQ---KLFARKMKILGTQILFLGIMTFSFG 65
Db 21 PAFEVL--EISPFQVSSGRLKLS--ASSPPLHTWLVTKQEFLGVQILTANICICFG 76
QY 66 VIFLFTLLKPYRPPFI--FLSGYPFGWSVLFGINSAPFLIAVKRKTTETLILSRINFL 123
Db 77 TVCVSLVDISHIEGDFSSFRAGYPFGWAIFFSIGMLSIISERRNATYLVKRGSGIGANTA 136
QY 124 SALGAIAGIILLTGFILDQNVICGYSHONS----QCKAVTVVLGLTLITLMTF-SIIE 177
Db 137 SSIAGTGITIL----IINLKKSLAYIIHSCQKFFETKCFMASFSTEIVVMMLFLTILG 192
QY 178 LFIISLPFSILGCHSE 192
Db 193 LGSVSLTICGAGEE 207

RESULT 6
TCR_STRAG
ID TCR_STRAG STANDARD; PRT; 458 AA.
AC P13924;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Tetracycline resistance protein.
GN TET.
OS Streptococcus agalactiae.
OC Plasmid pW158.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1311;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90016790; PubMed=2677995;
RA van der Lelle D., Bron S., Venema G., Oskam L.;
RT "Similarity of minus origins of replication and flanking open reading
RT frames of plasmids pUB110, pTB913 and pW158.";
RL Nucleic Acids Res. 17:7283-7294(1989).
CC [-] FUNCTION: RESISTANCE TO TETRACYCLINE BY AN ACTIVE TETRACYCLINE
CC EFFLUX. THIS IS AN ENERGY-DEPENDENT PROCESS THAT DECREASES THE

```


CC ACCUMULATION OF THE ANTIBIOTIC IN WHOLE CELLS. THIS PROTEIN
CC FUNCTIONS AS A METAL-TETRACYCLINE/H+ ANTI-PORTER.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN
CC AS THE DRUG RESISTANCE TRANSLOCASE FAMILY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X15669; CAA33712.1; -
DR PIR; C25599; YTSOG.
DR InterPro; IPR001411; TCR_TetB.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR PRINTS; PR01036; TCR_TetB.
KW Antibiotic resistance; Transmembrane; Transport; Symport; Plasmid.
FT TRANSMEM 12 33 POTENTIAL.
FT TRANSMEM 81 100 POTENTIAL.
FT TRANSMEM 111 129 POTENTIAL.
FT TRANSMEM 140 162 POTENTIAL.
FT TRANSMEM 165 185 POTENTIAL.
FT TRANSMEM 201 221 POTENTIAL.
FT TRANSMEM 223 240 POTENTIAL.
FT TRANSMEM 256 276 POTENTIAL.
FT TRANSMEM 297 317 POTENTIAL.
FT TRANSMEM 324 344 POTENTIAL.
FT TRANSMEM 346 365 POTENTIAL.
FT TRANSMEM 432 451 POTENTIAL.
SQ SEQUENCE 458 AA; 50006 MW; AD2014E7CA199995 CRC64;

Query Match 9.6%; Score 98.5; DB 1; Length 458;
Best Local Similarity 25.9%; Pred. No. 0.42;
Matches 45; Conservative 26; Mismatches 68; Indels 35; Gaps 7;
QY 32 TFSTQSLQKLFARKMKILGTIQLFGIMTFSGVIFLFTLLKPYPRPFILSGYPWG 91
DB 178 TIITVPFLMKLLKVRKRGHDK-GILMSVGIVFF-----MLFTSYSI-- 223
QY 92 SYLFINSGLAFIAVK--RKTETTLILSRIMNLSALGAIAGIILL--TFGILDQNYIC 147
DB 224 SFLIVSVLSFLFVHKIRKVDPFVDVGLGKNILFMIGVLCGGIIFGVAGFVSMVPTM 283
QY 148 GYSHNSOCKAVTVLFL-----GILL-----TLMTFSIIELFISLPF 184
DB 284 KDVLHOLSTAEIGSVITFFPTMSVIIFGYIGILVDRGRGLVXNLNIGVTLFSVF 337

RESULT 7
Y917_AQUAE
ID Y917_AQUAE STANDARD; PRT; 408 AA.
AC O67062;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AQ_917.
GN AQ_917.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., AuJay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex

RT aeolicus.";
RL Nature 392:353-358(1998).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; A5000713; AAC07022.1; -
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 15 37 POTENTIAL.
FT TRANSMEM 50 72 POTENTIAL.
FT TRANSMEM 87 104 POTENTIAL.
FT TRANSMEM 111 130 POTENTIAL.
FT TRANSMEM 145 167 POTENTIAL.
FT TRANSMEM 180 202 POTENTIAL.
FT TRANSMEM 236 258 POTENTIAL.
FT TRANSMEM 270 289 POTENTIAL.
FT TRANSMEM 304 326 POTENTIAL.
FT TRANSMEM 347 369 POTENTIAL.
FT TRANSMEM 374 393 POTENTIAL.
SQ SEQUENCE 408 AA; 47157 MW; F096D2EEB76BC022 CRC64;

Query Match 9.2%; Score 94; DB 1; Length 408;
Best Local Similarity 25.6%; Pred. No. 0.88;
Matches 52; Conservative 24; Mismatches 43; Indels 84; Gaps 11;
QY 36 QSPLOK-LFARKMKILGTIQLFGIMTFSGVIFLFTLLKPYPRPFILSGYPEW 90
DB 79 EDPLKATFVHIGLTLGLVYGLYSFNKVALLGFLLI----- 119
QY 91 GSVLFINSGLAFIAVKRKTETTLILSRIMN-----FLSALGAI-AGIILT----- 136
DB 120 -SLYITISS---LMLYK-----LLRIKRLDAPRGFRFALGSGTAGIILATFLVLNL 166
QY 137 FGFT-LDQNY-----ICGYSHQ-----NSQCKAVTVLF 163
DB 167 FGFNLNKKYLFDLHMSFMLFGWTATLVASVQVIEMFVTPPYPKFISSYLPKIVFTL 226
QY 164 LGLITLMTFSIIELFISLPFSI 186
DB 227 LVLKVFPLNFSLIDVFSILFTI 249

RESULT 8
TCR_BACST
ID TCR_BACST STANDARD; PRT; 458 AA.
AC P07561;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Tetracycline resistance protein.
GN TET.
OS Bacillus stearothermophilus.
OC Plasmid pTH15.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Geobacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86031344; PubMed=2996983;
RA Hoshino T., Ikeda T., Tomizuka N., Furukawa K.;
RT "Nucleotide sequence of the tetracycline resistance gene of pTH15, a
RT thermophilic Bacillus plasmid: comparison with staphylococcal TCR
RL controls.";
RL Gene 37:131-138(1985).
CC -1- FUNCTION: RESISTANCE TO TETRACYCLINE BY AN ACTIVE TETRACYCLINE
CC EFFLUX. THIS IS AN ENERGY-DEPENDENT PROCESS THAT DECREASES THE


```

DR InterPro: IPR000260; Oxidored_q5_N.
DR Pfam: PF00361; oxidored_q1; 1.
DR Pfam: PF01059; oxidored_q5_N; 1.
DR PRINTS: PR01437; NUOXDRDASE4.
DR KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
FT CONFLICT 161 161 S -> V (IN REF. 1; AAC47819).
SQ SEQUENCE 446 AA; 51371 MW; 8AB4A98563238BB2 CRC64;

Query Match      8.8%; Score 90; DB 1; Length 446;
Best Local Similarity 23.6%; Pred. No. 2;
Matches 34; Conservative 25; Mismatches 45; Indels 40; Gaps 6;

QY 66 VIFLTLKPYRPFIFSLGYPFW--GSVLFINSGAFLIAVRKKTTETLIILSRIMNFL 123
   : || |:: |||:: |||:: |||:: |||:: |||:: |||:: |||:: |||::
DB 5 IFLLFLI-----PFCFINNN-YMWVQIMMFFISIFL-----LMNNFMNTW 45
   : |||:: |||:: |||:: |||:: |||:: |||:: |||:: |||:: |||::
QY 124 SALGATAGAILTFFGLIDNVQICY-----SHNSQCKAVTVLFILGITLTMTFS 174
   : |||:: |||:: |||:: |||:: |||:: |||:: |||:: |||:: |||::
DB 46 SEISYFLGDMLSYGILLSLWCISMLLAASEMKNNYNKNFLNLIIILLLLLLTFS 105
   : |||:: |||:: |||:: |||:: |||:: |||:: |||:: |||:: |||::
QY 175 IELF-----ISLPFSILG 188
   : || | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 106 SMSLFMYLFFRESSLIPTLFILG 129
   : || | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 11
ID YAN9_SCHPO STANDARD; PRT; 656 AA.
AC Q10074;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 30-May-2000 (Rel. 39, Last annotation update)
DE Hypothetical 73.1 kDa protein C3H1.09C in chromosome I.
GN SPAC3H1.09C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OC NCBI_TaxId=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Gentles S., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
CC Submitted (DEC-1995) to the EMBL/GenBank/DDBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC CC -1- SIMILARITY: TO YEAST YKL146W AND YNL101W.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement or send an email to license@isb-sib.ch).
CC CC -----
DR EMBL; Z68144; CAAG2262.1; -
DR InterPro: IPR002422; AA_rel_permease_2.
DR Pfam: PF01490; Aa_trans; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 280 300 POTENTIAL.
FT TRANSEM 307 327 POTENTIAL.
FT TRANSEM 351 371 POTENTIAL.
FT TRANSEM 389 409 POTENTIAL.
FT TRANSEM 419 439 POTENTIAL.
FT TRANSEM 457 477 POTENTIAL.
FT TRANSEM 494 514 POTENTIAL.
FT TRANSEM 537 557 POTENTIAL.
FT TRANSEM 578 598 POTENTIAL.
FT TRANSEM 601 621 POTENTIAL.
FT TRANSEM 636 656 POTENTIAL.
SQ SEQUENCE 656 AA; 73060 MW; 7FD1E8F649ED83D2 CRC64;

```

```
Query Match      8.5%; Score 87.5; DB 1; Length 656;
Best Local Similarity 22.7%; Pred. NO. 4.6;
Matches 54; Conservative 34; Mismatches 75; Indels 75; Gaps 10;

QY 11 FLVPPPEITASEVESLSTFTSQSPLOKLFARKMKILGTIQLF-----57
DB 400 FLVFPV--LSLVKRSKLSATA-----LIADVILLGLLYFYFDVITLTKGIAD 448
QY 58 -----GIMTFSFGVIEFL-----TLKP--YPRPFELLSGYFPWGSVLPLIN 97
DB 449 VAMENKTDLSLFGVAITFYEGICILILPQEQMAKPNLPK-----LLGVMAAISLLFIS 504
QY 98 SCAFLIAVKRKTEFLIILSR-----IMNPLSALGAIAGIILTLFG--FILQNYICG 148
DB 505 IGLLSYAAGSGKVKVTWVILNMPSTFTVIIQFLYAIAILLSTPLQLPFAIAIEQIGFTR 564
QY 149 YSHQNSQCK-----AVTVLFGIILTLMTSITLFLSLPFLSGHSEDCDCEQC 200
DB 565 SGRNRKIKWRKNYLRVLIVLAILISWAGSRDLFLVSMGVS-----CC 610

RESULT 12
MTN2_MOUSE
ID MTN2_MOUSE STANDARD; PRT; 956 AA.
AC O08746;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE Matrilin-2 precursor.
GN MATN2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=Limb;
RX MEDLINE=97238863; Pubmed=9083061;
RA Deak F., Piecha D., Bachrati C., Paulsson M., Kiss I.;
RT Primary structure and expression of matrilin-2, the closest relative
RT of cartilage matrix protein within the von Willebrand factor type A-
RT like module superfamily.
RL J. Biol. Chem. 272:9268-9274 (1997).
CC -1- FUNCTION: INVOLVED IN MATRIX ASSEMBLY (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: DETECTED IN A VARIETY OF ORGANS, INCLUDING
CC CALVARIA, UTERUS, HEART, AND BRAIN, AS WELL AS FIBROBLAST AND
CC OSTEOBLAST CELL LINES.
CC -1- SIMILARITY: CONTAINS 10 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 VWFA DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U69262; AAC53163.1; -.
DR HSSP; P35555; IEMN.
DR MGD; MGI:109613; Matn2.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF-Ca.
DR InterPro; IPR002035; VWFA.
DR Pfam; PF00008; EGF; 10.
DR Pfam; PF00092; vwa; 2.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00181; EGF; 10.
DR SMART; SM00327; VWA; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 7.
```

```
DR PROSITE; PS01186; EGF_2; 9.
DR PROSITE; PS50234; VWFA_2.
KW EGF-like domain; Signal; Glycoprotein; Repeat; Coiled coil.
FT SIGNAL 1
FT CHAIN 24 956
FT DOMAIN 57 232 MATRILIN-2.
FT DOMAIN 238 278 EGF-LIKE 1.
FT DOMAIN 279 319 EGF-LIKE 2.
FT DOMAIN 320 360 EGF-LIKE 3.
FT DOMAIN 361 401 EGF-LIKE 4.
FT DOMAIN 402 442 EGF-LIKE 5.
FT DOMAIN 443 483 EGF-LIKE 6.
FT DOMAIN 484 524 EGF-LIKE 7.
FT DOMAIN 525 565 EGF-LIKE 8.
FT DOMAIN 566 606 EGF-LIKE 9.
FT DOMAIN 607 647 EGF-LIKE 10.
FT DOMAIN 655 830 VWFA 2.
FT DOMAIN 917 955 COILED COIL (POTENTIAL).
FT DISULFID 242 253 BY SIMILARITY.
FT DISULFID 249 262 BY SIMILARITY.
FT DISULFID 264 277 BY SIMILARITY.
FT DISULFID 283 294 BY SIMILARITY.
FT DISULFID 290 303 BY SIMILARITY.
FT DISULFID 305 318 BY SIMILARITY.
FT DISULFID 324 335 BY SIMILARITY.
FT DISULFID 331 344 BY SIMILARITY.
FT DISULFID 346 359 BY SIMILARITY.
FT DISULFID 365 376 BY SIMILARITY.
FT DISULFID 372 385 BY SIMILARITY.
FT DISULFID 387 400 BY SIMILARITY.
FT DISULFID 406 417 BY SIMILARITY.
FT DISULFID 413 426 BY SIMILARITY.
FT DISULFID 428 441 BY SIMILARITY.
FT DISULFID 447 458 BY SIMILARITY.
FT DISULFID 454 467 BY SIMILARITY.
FT DISULFID 469 482 BY SIMILARITY.
FT DISULFID 488 499 BY SIMILARITY.
FT DISULFID 495 508 BY SIMILARITY.
FT DISULFID 510 523 BY SIMILARITY.
FT DISULFID 529 540 BY SIMILARITY.
FT DISULFID 536 549 BY SIMILARITY.
FT DISULFID 551 564 BY SIMILARITY.
FT DISULFID 570 581 BY SIMILARITY.
FT DISULFID 577 590 BY SIMILARITY.
FT DISULFID 592 605 BY SIMILARITY.
FT DISULFID 611 622 BY SIMILARITY.
FT DISULFID 618 631 BY SIMILARITY.
FT DISULFID 633 646 BY SIMILARITY.
FT CARBOHYD 221 221 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 890 890 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 956 AA; 106779 MW; 3E4C22770B6E6EE CRC64;
```

```
Query Match      8.5%; Score 87.5; DB 1; Length 956;
Best Local Similarity 23.2%; Pred. NO. 6.4;
Matches 55; Conservative 31; Mismatches 76; Indels 75; Gaps 13;

QY 11 FLVPPPEIT--ASEYEST--ELSATFTSQSPLOKLFARKMKILGTIQLFQIMT---61
DB 86 FLDIGPDVTRVGLLYGSGTVKNEFSLTKFKKSEVERA-VKMRHLST-----GMTGLA 139
QY 62 --FSFGVIF-----LFTLLKPYPRPFFFLSGYPFWGSVLFINSGAFLIAVKRKTTETLI 114
DB 140 IQYALNIAFSEAEGARPLRNVPRIIMIVTDGRP-----QDSVAEVAAKARNTGLI 191
QY 115 ILSRI-----MNFLSALGA--TAGIILLFGIILQNYICGYSHQNSQCKAVTVLFLGLI 168
DB 192 FAIGVGQVDLNTLKAIGSEPHKDHVFLVNF-----SQIESLTSYFQNKLC 237
QY 169 TLMTFSIIE-----LFTSLPFS-----ILG-----CHSEDCDCEQC 200
DB 238 TVHMCVLEHNCALHCLNTPGSIYCKCKQGYILSTDQKTCRIQDLCDATEDHGCEQLC 294
```

```

RESULT 13
NUZM_APILI
ID NUZM_APILI STANDARD; PRN; 333 AA.
P34849; P92496; P92497; P92502; P92886; P92907; Q33785; Q33786;
Q33787; Q33788; Q36885;
01-FEB-1994 (Rel. 28, Created)
01-FEB-1994 (Rel. 28, Last sequence update)
30-MAY-2000 (Rel. 39, Last annotation update)
NADH-ubiquinone oxidoreductase chain 2 (EC 1.6.5.3).
ND2.
Apis mellifera ligustica (Common honeybee).
OS Mitochondrion.
OG Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;
OC Apoidea; Apidae; Apis.
ON NCBI_TaxID=7469;
OX [1]
XP SEQUENCE FROM N.A.
RP SEQUENCE OF 1-211 FROM N.A.
RC TISSUE=thorax;
RC MEDLINE=93114603; PubMed=8417993;
RA Crozier R.H., Crozier Y.C.;
RA "The mitochondrial genome of the honeybee Apis mellifera: complete
RT sequence and genome organization.";
RT Genetics 133:97-117(1993).
RN [2]
RP SEQUENCE OF 1-211 FROM N.A.
RC STRAIN-VARIOUS STRAINS;
RC MEDLINE=96364025; PubMed=8744768;
RA Arias M.C., Sheppard W.S.;
RA "Molecular phylogenetics of honey bee subspecies (Apis mellifera L.)
RT inferred from mitochondrial DNA sequence.";
RT Mol. Phylogenet. Evol. 5:557-566(1996).
RN [3]
RP SEQUENCE OF 60-307 FROM N.A.
RC STRAIN-HAPLOYPES 1 TO 11;
RA Koulianos S., Crozier R.H.;
RA Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RL -! CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -! SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
CC inner membrane.
CC
CC -! SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 2 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC
CC EMBL; L06178; AAB96798.1; -
CC EMBL; U35743; AAB38223.1; -
CC EMBL; U35744; AAB38224.1; -
CC EMBL; U35745; AAB38225.1; -
CC EMBL; U35746; AAB38226.1; -
CC EMBL; U35747; AAB38227.1; -
CC EMBL; U35748; AAD12744.1; -
CC EMBL; U35749; AAD12745.1; -
CC EMBL; U35750; AAB38228.1; -
CC EMBL; U35751; AAB38229.1; -
CC EMBL; U35752; AAB38230.1; -
CC EMBL; U35753; AAB38231.1; -
CC EMBL; U35754; AAB38232.1; -
CC EMBL; U35755; AAB38233.1; -
CC EMBL; U35756; AAB38234.1; -
CC EMBL; U35757; AAB38235.1; -
CC EMBL; U35758; AAB38236.1; -
CC EMBL; U35759; AAB38237.1; -
CC EMBL; U35760; AAB38238.1; -
CC EMBL; U35761; AAB38239.1; -
CC EMBL; U35762; AAB38241.1; -
CC EMBL; U35763; AAB38242.1; -

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 27, 2002, 11:02:16 ; Search time 19.84 Seconds
(without alignments)
968.642 Million cell updates/sec

Title: US-09-821-821-2

Perfect score: 1027

Sequence: 1 MDSSTAHSPVFLVFPPEITA.....SLPFSILGCHSDCDEQCQC 200

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR71.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1027	100.0	200	2	JC7585
2	181.5	17.7	214	2	I59258
3	173	16.8	291	2	A30558
4	169	16.5	297	1	A30586
5	158	15.4	243	2	A31231
6	157.5	15.3	235	2	B34342
7	123.5	12.0	244	2	A42806
8	106	10.3	307	2	C96844
9	103	10.0	455	2	T16070
10	100	9.7	152	2	JH0751
11	96	9.3	387	2	E81306
12	95.5	9.3	537	2	T07892
13	95	9.3	165	2	F69208
14	94.5	9.2	991	2	B71315
15	94	9.2	94	2	C70379
16	94	9.2	439	2	E95155
17	94	9.2	439	2	H98021
18	93.5	9.1	542	2	T01124
19	92.5	9.0	533	2	T07894
20	91.5	8.9	458	1	YTB506
21	91.5	8.9	458	1	YTB506
22	91.5	8.9	458	1	YTSOG
23	91.5	8.9	458	1	YTSOG
24	91	8.9	288	2	J01211
25	91	8.9	1431	2	A05235
26	90.5	8.8	457	2	T22748
27	90	8.8	446	2	T37205
28	89.5	8.7	1400	2	A71187
29	89	8.7	244	2	A71514
					H70193

30	89	8.7	353	2	T42971
31	88.5	8.6	275	2	H96981
32	88.5	8.6	538	2	T05714
33	88.5	8.6	538	2	T07164
34	88.5	8.6	540	2	T07604
35	88.5	8.6	694	2	E69143
36	88	8.6	485	2	E98261
37	88	8.6	485	2	AD3023
38	87.5	8.5	319	2	H69160
39	87.5	8.5	369	2	H90587
40	87.5	8.5	448	2	H97008
41	87.5	8.5	656	2	T38741
42	87.5	8.5	1227	2	T20370
43	87.5	8.5	1400	2	A81672
44	87	8.5	141	2	H89831
45	87	8.5	226	2	F84086

hypothetical prote
sugar permease [im
probable inorganic
phosphate transpor
hypothetical prote
L-asparagine perme
L-asparagine perme
hypothetical prote
hypothetical prote
probable cation ef
major facilitator
hypothetical prote
secDF protein, pro
hypothetical prote
hypothetical prote

ALIGNMENTS

RESULT 1
JC7585
testis expressed transmembrane-4 protein, TETM4 - human
C:Species: Homo sapiens (man)
C:Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 30-Jun-2001
C:Accession: JC7585
R:Hulett, M.D.; Pagler, E.; Hornby, J.R.; Hogarth, P.M.; Eyre, H.J.; Baker, E.; Crawford
Biochem. Biophys. Res. Commun. 280, 374-379, 2001
A:Title: Isolation, tissue distribution, and chromosomal localization of a novel test
A:Reference number: JC7585; MUID:21092614; PMID:11162526
A:Accession: JC7585
A:Molecule type: mRNA
A:Residues: 1-200 <HUL>
A:Cross-references: GB:AF321127
C:Comment: This protein, a four-transmembrane protein, associates with receptor compl
C:Genetics:
A:Gene: tetm4
A:Map position: 11q12
C:Keywords: signal transduction
F:1-48/Domain: cytoplasmic #status predicted <CYT1>
F:49-70/Domain: transmembrane #status predicted <TM1>
F:71-84/Domain: extracellular #status predicted <EXL1>
F:85-105/Domain: transmembrane #status predicted <TM2>
F:106-118/Domain: intracellular #status predicted <INT>
F:119-138/Domain: transmembrane #status predicted <TM3>
F:139-160/Domain: extracellular #status predicted <EXL2>
F:161-182/Domain: transmembrane #status predicted <TM4>
F:183-200/Domain: cytoplasmic #status predicted <CYT2>

Query Match 100.0%; Score 1027; DB 2; Length 200;
Best Local Similarity 100.0%; Pred. No. 1.5e-87;
Matches 200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDSSTAHSPVFLVFPPEITA SEYESTELSATSTESQPLQKLFARKMKILGTITQILFGIM 60
Db 1 MDSSTAHSPVFLVFPPEITA SEYESTELSATSTESQPLQKLFARKMKILGTITQILFGIM 60
QY 61 TFSFGVIFLTLKPPRPFPIELSGYPFWGSLVFNINSAGFLIAVRKKTETLIILSRIM 120
Db 61 TFSFGVIFLTLKPPRPFPIELSGYPFWGSLVFNINSAGFLIAVRKKTETLIILSRIM 120
QY 121 NFSLALGAIAIGIILLTFGLTDQNYICGYSHQNSQCKAVTVLFGILITLMTFSIELFI 180
Db 121 NFSLALGAIAIGIILLTFGLTDQNYICGYSHQNSQCKAVTVLFGILITLMTFSIELFI 180
QY 181 SLPFSILGCHSDCDEQCQC 200
Db 181 SLPFSILGCHSDCDEQCQC 200

RESULT 2


```
Db 36 SLVGPQSFMRBRSKVLGAVQIMNGLFHIALGGL-----LMIPAGIYAPICVTWYPLWGG 91
Qy 93 VLFINSAGFLIYAKRKTETLTIILSRIMNFLSALGAIGIILL-----T 136
Db 92 IMYIISGLAATEKRSKRLVKGMKIMNSLSIFAISGMILSMDILNIKISHFLKMES 151
Qy 137 FGIILQON-YICGY-----SHONS-----QCKAVTVLFLGILITLMTFSII-ELFIS 181
Db 152 LNFIRAHPTPIYINYNCEPANPSEKNSPSTQYCYISQSLFGLILSVMLIFAFQOELVIA 209

RESULT 5
A42331
Ige Fc receptor beta chain - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 05-Nov-1999
C:Accession: A31231
R:Kinet, J.P.; Blank, U.; Ra, C.; White, K.; Metzger, H.; Kochan, J.
Proc. Natl. Acad. Sci. U.S.A. 85, 6483-6487, 1988
A:Title: Isolation and characterization of cDNAs coding for the beta-subunit of the high
A:Reference number: A42331; MUID:88320465
A:Accession: A31231
A:Molecule type: mRNA
A:Residues: 1-243 <KIN>
A:Cross-references: GB:M22923; GB:J03845; NID:g204116; PIDN:AAA41149.1; PID:g204117
C:Keywords: immunoglobulin receptor; transmembrane protein

Query Match 15.4%; Score 158; DB 2; Length 243;
Best Local Similarity 28.0%; Pred. No. 2.9e-07;
Matches 45; Conservative 28; Mismatches 72; Indels 16; Gaps 3;

Qy 36 QSPQKLFARKMKILGTIQLFGITFSGVIFLFTLLKP--YPRFFIFLSGVPFGSV 93
Db 47 QQTWQSFLKKELEFGLVQVLVGLICLCFGFVVCSTIQTSDFDDEVLLVIRAGYPFWGAV 106

Qy 94 LFINSGAFLIAVAKRKTETLTIILSRIMNFLSALGAIGIILLTFPGFILDQNYICGYSHQN 153
Db 107 LEVLGSLIMSERKNTLYLVRSGLGANIVSSIAAGLIAITLNLNNSAYM----- 159

Qy 154 SOCKAVT-----VLFGLITLMTFSIIEFLFSLPSIL 187
Db 160 NYCKDITEDDGCFTVSTFTELVLMLLFTLILAFCSAVILLII 200

RESULT 6
B34342
Ige Fc receptor beta chain - mouse
C:Species: Mus musculus (house mouse)
C>Date: 15-Jun-1990 #sequence_revision 15-Jun-1990 #text_change 05-Nov-1999
C:Accession: B34342
R:Ra, C.; Jouvin, M.H.E.; Kinet, J.P.
J. Biol. Chem. 264, 15323-15327, 1989
A:Title: Complete structure of the mouse mast cell receptor for IgE (Fc-epsilon-RI) and
A:Reference number: A34342; MUID:89359361
A:Accession: B34342
A:Molecule type: mRNA
A:Status: preliminary
A:Residues: 1-235 <RAC>
A:Cross-references: GB:J05019; NID:g193238; PIDN:AAA37601.1; PID:g309225
C:Keywords: immunoglobulin receptor; transmembrane protein

Query Match 15.3%; Score 157.5; DB 2; Length 235;
Best Local Similarity 30.5%; Pred. No. 3.1e-07;
Matches 60; Conservative 28; Mismatches 82; Indels 27; Gaps 9;

Qy 1 MDSSTAHPVFLVFP-PEITASVESTELSATFTSQSPLOQ--KLPARK-MKILGTIQL 56
Db 1 MDTEN-RSRADLALPNQESSAPDTELLEASPAKAPKQWRTPLKKEFLFGATQIL 59

Qy 57 FGIWTFSGVIFLFTLLKPY-----PRFFIFLSGYPFGWSVLFINSGAFLIAVAKRKTTE 111
```

```
Db 60 VGLICLCFGTIVCSVL---YVSDFDEEVLLLYKLGYFWGAVLFLVSGFLSIISERKNTL 116
Qy 112 TLTIILSRIMNFLSALGAIGIILLTFGLFDQNYICGYSHONSQCKAVT-----VLFL 164
Db 117 YLVRSGLANIVSSIAAGTGAMILL-----NLTNFAYMNN-CKNVTEDDGCVFASFT 169
Qy 165 GILITLMTFSIIEFLIS 181
Db 170 TELVLMMLFLTLAFCS 186

RESULT 7
A42806
Ige Fc receptor beta chain - human
C:Species: Homo sapiens (man)
C>Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 08-Oct-1999
C:Accession: A42806; S21154
R:Kuester, H.; Zhang, L.; Brini, A.T.; MacGlashan, D.W.J.; Kinet, J.P.
J. Biol. Chem. 267, 12782-12787, 1992
A:Title: The gene and cDNA for the human high affinity immunoglobulin E receptor beta
A:Reference number: A42806; MUID:92316966
A:Accession: A42806
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-244 <KUE>
A:Cross-references: GB:M89796; NID:g337417; PIDN:AAA60269.1; PID:g337418
R:Maekawa, K.; Imagawa, N.; Tanaka, Y.; Harada, S.
FEBS Lett. 302, 161-165, 1992
A:Title: Determination of the sequence coding for the beta subunit of the human high-
A:Reference number: S21154; MUID:92339505
A:Accession: S21154
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-244 <MAE>
A:Cross-references: GB:D10583; NID:g219881; PIDN:BAAO1440.1; PID:d1001914; PID:g21988
C:Genetics: 19/2; 62/3; 107/3; 126/3; 179/3; 179/3; 212/3
A:Introns: 19/2; 62/3; 107/3; 126/3; 179/3; 179/3; 212/3
C:Keywords: immunoglobulin receptor; transmembrane protein

Query Match 12.0%; Score 123.5; DB 2; Length 244;
Best Local Similarity 25.6%; Pred. No. 0.00044;
Matches 50; Conservative 33; Mismatches 93; Indels 19; Gaps 7;

Qy 9 PVFLVPPPEITASEYESTELSATFTSQSPLOQ---KLPARKMKILGTIQLFGIMTFSG 65
Db 21 PAFEVL--EISPOEVSSGRLLKS--ASSPPLHTWLTVLKKEQEFGLVQTLTAMICLCFG 76

Qy 66 VIFLFTLLKPYPRPPEI--FLSGYPFGWSVLFINSGAFLIAVAKRKTETLTIILSRIMNFL 123
Db 77 TVVCSVLDSIHIEDTFSSFKAGYPFWGAIFFSISGMLSIISERNATYLVRSGLGANTA 136

Qy 124 SALGAIGIILLTFGLFDQNYICGYSHONS-----QCKAVTVLFLGILITLMTF-SITE 177
Db 137 SSIAGG*GITIL----IINLKKSLAYIIHSCORFFETKCFMASFSTEIVYVMMFLTILG 192

Qy 178 LFIISLPFSILGCHSE 192
Db 193 LGSVAVSUTICGAGEE 207

RESULT 8
C86844
Hypothetical protein ysbC [imported] - Lactococcus lactis subsp. lactis (strain IL140
C:Species: Lactococcus lactis subsp. lactis
C>Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: C86844
R:Boilotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Eh
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: C86844
A:Status: preliminary
```

A:Molecule type: DNA
A:Residues: 1-307 <STO>
A:Cross-references: GB:AE005176; PID:g12724776; PIDN:AAK05853.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: ysbC

Query Match 10.3%; Score 106; DB 2; Length 307;
Best Local Similarity 31.9%; Pred. No. 0.023;
Matches 46; Conservative 21; Mismatches 49; Indels 28; Gaps 9;

QY 34 STQSPLOKLFARKKILGTQILFGIMTSEFGVIEL--FTLLKPYPRP--FIFLSGYPPFW 90
DB 16 ANQPIN---ADLRKIVGSPFLASISNF--VGSIFIGIITLTSOTLPSPFQVGSHPAW 71

QY 91 ---GSVLFINSGAFIAVKRKTETLILSRIMNFSALGAIAAGIILTFGLDQNYIC 147
DB 72 IWIGGVL---GGIFL-----TSNVLFPFR-----LGAVQTVILPILGOILMGTLID 114

QY 148 GYSHONSOCRAVTVL-FLGILITL 170
DB 115 SFGWFHAMQLPMTLMRLGLIIITL 138

RESULT 9

Ti6070
hypothetical protein F14B8.3 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T16070

submitted to the EMBL Data Library, June 1995
A:Description: The sequence of *C. elegans* cosmid F14B8.

A:Reference number: Z18456

A:Accession: T16070

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-455 <GEI>

A:Cross-references: EMBL:U28737; NID:g860717; PID:g860719; PIDN:AAA68272.1; CESP:F14B8.3

A:Experimental source: strain Bristol N2

C:Genetics:

A:Gene: CESP:F14B8.3

A:Introns: 56/3; 193/2; 224/3; 289/2; 302/3; 370/2; 402/3; 442/3

Query Match 10.0%; Score 103; DB 2; Length 455;
Best Local Similarity 31.6%; Pred. No. 0.063;
Matches 31; Conservative 15; Mismatches 32; Indels 20; Gaps 3;

QY 49 ILGTQILFGIMT-----FSFGVIFLETLLKPYPR--PPFIFLSGYPPFW 91
DB 303 LLGIVEYFGLISRNHLISICINGISMCAGSVLILFGLQKNYPTIYVFLIGY---A 359

QY 92 SVLFINSGAFIAVKRKTETLILSRIMNFSALGAI 129
DB 360 SIEFIWDACYLCVVEQVPTVETRGTSAGCSFLSLSGI 397

RESULT 10

JH0751
Ige receptor beta chain - human (fragment)

C:Species: *Homo sapiens* (man)

C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000

C:Accession: JH0751

R:Bieber, T.; de la Salle, H.; Wollenberg, A.; Hakimi, J.; Chizzonite, R.; Ring, J.; Han

J. Exp. Med. 175, 1285-1290, 1992

A:Title: Human epidermal Langerhans cells express the high affinity receptor for immunog

A:Reference number: JH0751; MUID:92235608

A:Accession: JH0751

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-152 <BIE>

A:Cross-references: GB:X66187; NID:g396463; PIDN:CAA46955.1; PID:g396464

Query Match 9.7%; Score 100; DB 2; Length 152;
Best Local Similarity 27.6%; Pred. No. 0.042;
Matches 32; Conservative 21; Mismatches 53; Indels 10; Gaps 3;

QY 83 FLUSGYPFWGSLVFINSGAFIAVKRKTETLILSRIMNFSALGAIAAGIILTFGLD 142
DB 26 FRAGYFPWGAFFSISGMSIISERRNATYVRGSLGANTASSIAGGTGITIL---IIN 81

QY 143 QNYICGYSHONS-----QCKAVTVLFLGILITLMTF--SIIELFISLPSILGCHSE 192
DB 82 LKSLAYIIHHSCKFFETKCFMASFSTEIVVMMLFTLILGLSAVSULTICGAGEE 137

RESULT 11

EB1306
probable transmembrane transport protein Cj1040c [imported] - *Campylobacter jejuni* (s
C:Species: *Campylobacter jejuni*
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Mar-2000
C:Accession: EB1306

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chil
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Ba
Nature 403, 665-668, 2000

A:Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals

A:Reference number: A81250; MUID:20150912

A:Accession: EB1306

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-387 <PAR>

A:Cross-references: GB:AL139077; GB:AL111168; NID:g6968444; PIDN:CAB73296.1; PID:g696

A:Experimental source: serotype O2, strain NCCT 11168

C:Genetics:

A:Gene: Cj1040c

Query Match 9.3%; Score 96; DB 2; Length 387;
Best Local Similarity 25.0%; Pred. No. 0.24;
Matches 42; Conservative 31; Mismatches 65; Indels 30; Gaps 9;

QY 14 PPEITASEYESTELSATTFSTOSQLOKLPARKKILGTIOI--LFGIMTSEFGVIFLFT 71
DB 233 FDKFEFSWLVFAQLVAAPVSLFGPL--LLGKURQNLHFFYIAGLCSWVYIAFGMLFIFD 290

QY 72 LKPYPRPFI---FLSGYPFWGSLVFINSGAFIAVKRKTETLILSRIMNFSALGA 128
DB 291 -----SKTIIISAFIMGFP--WGVF----GIALLFIAQKSSN-----AQIAARLSALAQ 335

QY 129 TAGIITLTFGLDQNYICGYSH--ONSCKAVTVLFLGILITLMTF 173

DB 336 GFGYLIAAQ-----QWIIIGLHKPFNSFALLMLVFGVILNIFGY 378

RESULT 12

T07892

probable inorganic phosphate transport protein Ptl - barrel medic

C:Species: *Medicago truncatula* (barrel medic)

C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 23-Mar-2001

C:Accession: T07892

R:Liu, H.; Trieu, A.T.; Blaylock, L.A.; Harrison, M.J.

Mol. Plant Microbe Interact. 11, 14-22, 1998

A:Title: Cloning and characterization of two phosphate transporters from *Medicago tru*

A:Reference number: Z16196; MUID:98086876

A:Accession: T07892

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-537 <LIU>

A:Cross-references: EMBL:AF000354; NID:g2465526; PIDN:AAB81346.1; PID:g2465527

C:Genetics:

A:Gene: Ptl

C:Superfamily: probable inorganic phosphate transport protein PHO84

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 27, 2002, 11:02:16 ; Search time 65.04 Seconds
(without alignments)
341.555 Million cell updates/sec

Title: US-09-821-821-2
Perfect score: 1027
Sequence: 1 MDSSTAHSFVFLVFPPEITA.....SLPFSILCHSEDCDCQCC 200

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*

```

1: /SIDS1/gcgdata/hold-geneseg/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/hold-geneseg/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/hold-geneseg/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/hold-geneseg/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/hold-geneseg/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/hold-geneseg/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/hold-geneseg/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/hold-geneseg/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/hold-geneseg/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/hold-geneseg/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/hold-geneseg/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/hold-geneseg/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/hold-geneseg/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/hold-geneseg/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/hold-geneseg/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/hold-geneseg/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/hold-geneseg/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/hold-geneseg/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/hold-geneseg/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/hold-geneseg/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/hold-geneseg/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/hold-geneseg/geneseq-emb1/AA2001.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1027	100.0	200	22	AAE13062 Human CD20/IgE-rec
2	1027	100.0	200	22	AAE13062 Human secreted pro
3	1004	97.8	199	22	AAU01210 Novel human membra
4	937	91.2	201	22	AAE10917 Human gene 9 encod
5	937	91.2	212	22	AAE05114 Human reproductive
6	735.5	71.6	149	22	AAU01213 Novel human membra
7	687	66.9	158	22	ABU12234 Human secreted pro
8	582	56.7	138	22	AAU01212 Novel human membra
9	414	40.3	81	22	AAU01211 Novel human membra
10	294	28.6	77	21	AAE03074 Human secreted pro
11	263.5	25.7	67	22	AAE03074 Human secreted pro

12	202.5	19.7	239	20	AAW96745	High affinity immu
13	202.5	19.7	239	21	AAW96745	Human high affinity
14	202.5	19.7	245	21	AAW94973	Human secreted pro
15	201.5	19.6	220	22	AAW93512	Human polypeptide,
16	183.5	17.9	225	21	AAW94449	Human inflammation
17	183.5	17.9	242	21	AAW73495	Human secreted pro
18	183.5	17.9	248	20	AAW15225	Human receptor pro
19	183.5	17.9	248	21	AAW91531	Human secreted pro
20	183.5	17.9	248	22	AAE12072	Dendritic cell (DC
21	183.5	17.9	248	22	AAE12072	Human hHAIRBs-iso
22	183.5	17.9	250	20	AAW48505	Human breast tumo
23	183.5	17.9	257	21	AAW58419	Lung cancer associ
24	183.5	17.9	273	21	AAW91680	Human secreted pro
25	183.5	17.9	273	22	ABU11989	Human secreted pro
26	183.5	17.9	273	22	AAW25809	Human protein sequ
27	183.5	17.9	273	22	AAW75619	Human colon cancer
28	182.5	17.8	225	22	AAE65272	Human cell surface
29	181.5	17.7	214	17	AAW06503	HTM4 protein. Hom
30	181.5	17.7	214	19	AAW41056	HTM4 protein. Hom
31	180	17.5	226	22	AAW98706	Chandra, a helper
32	176.5	17.2	227	22	ABG19236	Novel human diagno
33	172.5	16.8	267	20	AAW15224	Human receptor pro
34	172.5	16.8	273	22	ABG17004	Novel human diagno
35	171	16.7	178	22	AAE12073	Dendritic cell (DC
36	171	16.7	297	13	AAE20808	Human CD20 antigen
37	170.5	16.6	299	21	AAW91352	Human secreted pro
38	169	16.5	297	10	AAW91356	CD20.4 antigen.
39	169	16.5	297	21	AAW96131	Human cell surface
40	169	16.5	297	22	AAU02440	Human lymphocyte c
41	168	16.4	239	22	AAW74584	Human HATERBs isof
42	168	16.4	248	20	AAW36046	Extended human sec
43	168	16.4	249	22	AAW74583	Human HATERBs isof
44	168	16.4	297	17	AAW91436	Human CD20 antigen
45	168	16.4	297	19	AAW80445	Human CD20.4 antig

ALIGNMENTS

```

RESULT 1
AAE13062
ID AAE13062 standard; Protein; 200 AA.
AC AAE13062;
DT 28-JAN-2002 (first entry)
DE Human CD20/IgE-receptor like protein, agp-96614-al.
XX
XX Human; CD20/IgE-receptor like protein; immunoglobulin E; agp-96614-al;
XX agp-69406-al; cancer; abnormal cell proliferation; autoimmune disease;
XX ovarian cancer; brain cancer; arteriosclerosis; vascular restenosis;
XX rheumatoid arthritis; multiple sclerosis; allergy; dermatitis; asthma;
XX reproductiv disease; diabetes; diabetes; transplant rejection; endometriosis;
XX infertility; gene therapy.
XX
XX Homo sapiens.
XX
XX WO200174903-A2.
XX
XX 11-OCT-2001.
XX
XX 29-MAR-2001; 2001WO-US10048.
XX
XX 30-MAR-2000; 2000US-193728P.
XX
XX 27-NOV-2000; 2000US-0723258.
XX
XX (AMGE-) AMGEN INC.
XX
XX Welcher AA, Calzone FJ;
XX
XX WPI; 2001-662968/76.
XX
XX N-PSDB; AAD21441.

```

```

XX Novel CD20/IgE-receptor like polypeptides and polynucleotides,
PT antagonists and antibodies of the polypeptide useful for treating
PT ameliorating or preventing diseases associated with the polypeptide
PT e.g. cancer, asthma
XX
PS Claim 13; Fig 1; 145pp; English.
XX
CC The invention relates to human CD20/immunoglobulin E (IgE)-receptor
CC like polypeptides designated as agp-96614-al and agp-69406-al and
CC nucleic acid molecules encoding such polypeptides. Polypeptides of
CC the invention are useful for treating, preventing or ameliorating
CC a disease, condition, or disorder which includes cancer such as
CC brain cancer, ovarian cancer; abnormal cell proliferation such as
CC arteriosclerosis, vascular restenosis; pathology from allergens
CC such as allergies, asthma, dermatitis; dysfunction of immune system
CC such as rheumatoid arthritis, autoimmune disease, multiple sclerosis,
CC diabetes, transplant rejection and reproductive diseases such as
CC infertility, preterm labour and delivery, endometriosis etc. They
CC are also useful for identifying antagonists and as immunogens, for
CC raising antibodies which may also be used to prevent, treat or
CC diagnose a number of diseases and disorders. Polynucleotides of the
CC invention are used to map the location of CD20/IgE-receptor like
CC gene and related genes on chromosomes and as hybridisation probes.
CC They are also useful in gene therapy. The present sequence is
CC human CD20/IgE-receptor like protein, agp-96614-al.
XX
SQ Sequence 200 AA;

Query Match 100.0%; Score 1027; DB 22; Length 200;
Best Local Similarity 100.0%; Pred. No. 2.9e-111;
Matches 200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDSSTAHSPVFLVFPPEITASEYESTELSATFTSQSPLOKLFARKMKILGTIQLFGIM 60
DB 1 mdsstahspvflvfppeitaseyestelsattfstqspklfarkmkilgtiqlfgim 60

QY 61 TFSFGVIFLTLKPYPRPFIFLSGYPTWGSVLFINSAGFLIAVKRKTETLILSRIM 120
DB 61 tfsfgviflftllkpyprpfiflsgyptwgsvlfinsgafliavkrktetlilsrim 120

QY 121 NPLSALGAIAGIILLTFGFIIDQNYICGYSHQNSOCKAVTVLFLGLITLMTFSIIEIFI 180
DB 121 nplsalgaiagiilltfgfildqnyicyshqnsqckavtvflfgliltmtfsiieifi 180

QY 181 SLPPFSILGCHSEDCCEQCC 200
DB 181 slpfsilgchsedcdceqcc 200

RESULT 2
AAG89138 11-SEP-2001 (first entry)
ID AAG89138 standard; Protein; 200 AA.
XX AAG89138;
XX
DT 11-SEP-2001 (first entry)
XX
DE Human secreted protein, SEQ ID NO: 258.
XX
KW Human; secreted protein; gene therapy; vaccine; treatment; diagnosis;
KW GENSET.
XX
OS Homo sapiens.
XX
PN WO200142451-A2.
XX
PD 14-JUN-2001.
XX
PF 07-DEC-2000; 2000WO-IB01938.
XX
XX 08-DEC-1999; 99US-0169629.
XX

```

```

PR 06-MAR-2000; 2000US-0187470.
XX (GEST ) GENSET.
PA
XX Dumas Milne Edwards J, Bougueleret L, Jobert S;
FI WPI; 2001-367870/38.
DR N-PSDB; AAH64741.
XX
XX Full length GENSET human nucleic acids encoding potentially secreted
PT proteins, useful in gene therapy and vaccination against a variety of
PT diseases, and for diagnosis of those diseases -
XX Claim 21; Page 802-803; 921pp; English.
XX
CC The invention relates to full length GENSET human nucleic acids encoding
CC potentially secreted proteins. The nucleic acids and the polypeptides
CC they encode may be used in the prevention, treatment and diagnosis of
CC diseases associated with inappropriate GENSET gene expression. For
CC example, they be used to treat disorders associated with decreased
CC GENSET gene expression by rectifying mutations or deletions in a
CC patient's genome that affect the activity of GENSET or by supplementing
CC the patient's own production of GENSET polypeptides. Conversely,
CC antisense nucleic acid molecules may be administered to down regulate
CC GENSET expression by binding with the cells' own genes and preventing
CC their expression. The sense and antisense nucleic acids may also be
CC used as DNA probes in diagnostic assays to detect and quantitate the
CC presence of similar nucleic acid sequences in samples, and hence to
CC determine which patients may be in need of restorative therapy.
CC The GENSET polypeptides may be used as antigens in the production of
CC antibodies and in assays to identify modulators (agonists and
CC antagonists) of GENSET polypeptide expression and activity. The
CC present sequence is a GENSET polypeptide of the invention.
XX
SQ Sequence 200 AA;

Query Match 100.0%; Score 1027; DB 22; Length 200;
Best Local Similarity 100.0%; Pred. No. 2.9e-111;
Matches 200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDSSTAHSPVFLVFPPEITASEYESTELSATFTSQSPLOKLFARKMKILGTIQLFGIM 60
DB 1 mdsstahspvflvfppeitaseyestelsattfstqspklfarkmkilgtiqlfgim 60

QY 61 TFSFGVIFLTLKPYPRPFIFLSGYPTWGSVLFINSAGFLIAVKRKTETLILSRIM 120
DB 61 tfsfgviflftllkpyprpfiflsgyptwgsvlfinsgafliavkrktetlilsrim 120

QY 121 NPLSALGAIAGIILLTFGFIIDQNYICGYSHQNSOCKAVTVLFLGLITLMTFSIIEIFI 180
DB 121 nplsalgaiagiilltfgfildqnyicyshqnsqckavtvflfgliltmtfsiieifi 180

QY 181 SLPPFSILGCHSEDCCEQCC 200
DB 181 slpfsilgchsedcdceqcc 200

RESULT 3
AAU01210
ID AAU01210 standard; Protein; 199 AA.
XX
XX AAU01210;
XX
XX 26-SEP-2001 (first entry)
XX
XX Novel human membrane protein #1.
XX
XX Human; membrane protein; membrane receptor; IgE receptor; CD20;
KW physiological disorder.
XX
XX Homo sapiens.
XX

```



```

FH Key Location/Qualifiers
FT Misc-difference 199 /note= "Encoded by TGTGT"
PN WO200146417-A2.
XX 28-JUN-2001.
XX 12-DEC-2000; 2000WO-US33742.
XX 22-DEC-1999; 99US-0171567.
XX (LEXI-) LEXICON GENETICS INC.
XX Walke DW, Turner CA;
XX WPI; 2001-408646/43.
XX N-PSDB; AAS04279.
XX Polynucleotide encoding novel human membrane protein, useful for
FT identifying agonist, antagonist or modifiers or for producing
FT antibodies useful in therapeutic, diagnostic and pharmacogenomic
FT applications -
XX Claim 2; Page 29-30; 32pp; English.
XX The present sequence represents novel human membrane protein #1.
CC Human membrane protein #1 is 1 of 4 human membrane proteins
CC (AAU01210-AAU01213) given in the present invention. These membrane
CC proteins share structural similarity with membrane receptors such as
CC the 19c receptor and mammalian CD20. The novel human membrane proteins
CC are useful for identifying agonists, antagonists and modulators of the
CC membrane proteins, and for producing antibodies specific to the
CC membrane proteins. The membrane proteins can be used for diagnosis,
CC drug screening, pharmacogenomic applications, clinical trial monitoring
CC and the treatment of physiological disorders and diseases. The
CC polynucleotides encoding the membrane proteins can be used to generate
CC PCR primers or probes to identify mutations associated with a particular
CC disease.
XX Sequence 199 AA;
SQ
Query Match 97.8%; Score 1004; DB 22; Length 199;
Best Local Similarity 99.0%; Pred. No. 1.4e-108;
Matches 197; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 MDSSTAHSVPVLVFPPEITASEYESTELSATSTSTQSPLOKLFARKMKILGTIQLFGIM 60
Db 1 mdsstahspvflvppeitaseyestelsattstqspqlkfarkmklgtlqlflgim 60
Qy 61 TFSFGVIFLTLKPYRPFIFLSGYPFWGSLVFINSGAFLIAVRKRTTETLILSRIM 120
Db 61 tfsfgvifltlkpyrpfiflsgypfwgslvfinsgafliavrkrttetiilslrim 120
Qy 121 NFLSALGAIAGIILTFGFIIDQNYICGYSHQNSQCKAVTVLFGIILITLMTFSIIEFI 180
Db 121 nllsalragiilftfgfiidqnyicgyshqnsqckavtvlfgiilitlmtfsiiefi 180
Qy 181 SIPLFSILGCHSDCCEOC 199
Db 181 siplfsglchsdcdceqc 199
RESULT 4
AAE10917
ID AAE10917 standard; Protein; 201 AA.
XX AC AAE10917;
XX 18-DEC-2001 (first entry)
XX DE Human gene 9 encoded immune system-related protein HTENNA45.

```

```

XX Human; immune system-related protein; allergy; rheumatoid arthritis;
KW cancer; Gaucher's disease; viral hepatitis; gene therapy; cytostatic;
KW diabetes mellitus; arrhythmia; wound healing; ischaemic lesion; AIDS;
KW Acquired Immune Deficiency Syndrome; virucide; hepatotropic; vasotropic;
KW autoimmune disorder; inflammation; cardiovascular disorder; hair loss;
KW wound healing; cell proliferation; skin aging; endocrine disorder;
XX Homo sapiens.
XX Key Location/Qualifiers
FH Domain 21..26
FT /label= Immunogenic_epitope
FT Domain 150..156
FT /label= Immunogenic_epitope
XX WO200166722-A1.
XX 13-SEP-2001.
XX 07-MAR-2001; 2001WO-US07260.
XX 08-MAR-2000; 2000US-187873P.
XX 11-AUG-2000; 2000US-224367P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ni J, Hilbert D, Kenny JJ, Moore PA, Choi GH, Soppet DR, Ebner R;
XX Gruber JR, Endress GA, Ruben SM;
XX WPI; 2001-589939/66.
XX N-PSDB; AAD18275.
XX Novel isolated immune system-related polypeptide useful for treating
PT rheumatoid arthritis, AIDS, allergy, cancer, Gaucher's disease,
PT diabetes mellitus, arrhythmia, wound healing, ischemic lesions and
PT viral hepatitis -
XX Claim 11; Page 310-311; 315pp; English.
XX The invention relates to human immune system-related protein and their
CC DNA. Human immune-system related protein and DNA are useful for
CC preventing, treating or ameliorating a medical condition in a mammalian
CC subject, for diagnosing, preventing or treating immune system-associated
CC disorders, autoimmune disorders (rheumatoid arthritis), inflammatory
CC disorders (allergies), immunodeficiencies (AIDS), inflammatory disorders
CC (allergies), infectious diseases (e.g., viral hepatitis), complement
CC activation disorders, immune complex diseases, neoplastic disorders
CC (cancer), hyperproliferative disorders (Gaucher's disease), disorders
CC associated with neovascularisation, diseases at the cellular level,
CC cardiovascular disorders (arrhythmias), wound healing and epithelial
CC cell proliferation, endocrine disorders (diabetes mellitus) and
CC neurological disorders (ischaemic lesions). Immune-system related protein
CC or DNA is useful for preventing hair loss, skin aging due to sunburn, to
CC maintain organs before transplantation, to treat weight disorders, to
CC modulate mammalian characteristics, to change a mammal's mental or
CC physical state, or as a food additive or preservative. Immune-system
CC related DNA is useful in gene therapy, for chromosome identification,
CC radiation hybrid mapping, long range restriction mapping and in forensic
CC biolog. The present sequence represents a human immune-system related
CC protein of the invention.
XX Sequence 201 AA;
SQ
Query Match 91.2%; Score 937; DB 22; Length 201;
Best Local Similarity 99.5%; Pred. No. 8.8e-101;
Matches 187; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MDSSTAHSVPVLVFPPEITASEYESTELSATSTSTQSPLOKLFARKMKILGTIQLFGIM 60
Db 1 mdsstahspvflvppeitaseyestelsattstqspqlkfarkmklgtlqlflgim 60

```


PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-465570/50.
DR N-PSDB; AAL01084.
XX
XX Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition -
PT
XX
PS Claim 11; SEQ ID NO 3772; 1297pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a protein of the invention.
XX
XX Sequence 212 AA;
SQ

Query Match 91.2%; Score 937; DB 22; Length 212;
Best Local Similarity 99.5%; Pred. No. 9.4e-101;
Matches 187; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDSSTAHSPVFLVFPPEITASEYESTELSATTFSTQSPLOKLFARKMKILGTIQLFGIM 60
Db 12 mdsstahspvflvfppeitaseyestelsattfstqspqlkfarkmkilgtiqlfgim 71
QY 61 TFSFGVIFLTLKPYPRPFIFLSGYFPWGSVLFINSAGFLIAVKRKTETLLILSRIM 120
Db 72 tfsfgvifltlkpyprpfiflsgypfwgsvlfinsgafliavkrktetllilslrim 131
QY 121 NFLSALGATAGIILLFGFTLDQNYICGYSHQNSQCKAVTVLFLGILITLMTFSIIEIFI 180
Db 132 nflsalgalagaillltfgfildqnyicgyshqnsqckavtvflflgilitlmtfsiieifi 191
QY 181 SLPFSIIG 188
Db 192 slpfsiig 199

RESULT 6
AAU01213
ID AAU01213 standard; Protein; 149 AA.
XX
AC AAU01213;
XX

DT 26-SEP-2001 (first entry)
XX
DE Novel human membrane protein #4.
XX
KW Human; membrane protein; membrane receptor; IgE receptor; CD20;
KW Physiological disorder.
XX
OS Homo sapiens.
XX
PN WO200146417-A2.
XX
XX 28-JUN-2001.
XX
XX 12-DEC-2000; 2000WO-US33742.
XX
XX 22-DEC-1999; 99US-0171567.
XX (LEXI-) LEXICON GENETICS INC.
XX
XX Walke DW, Turner CA;
XX
XX WPI; 2001-408646/43.
DR N-PSDB; AAS04282.
XX
XX Polynucleotide encoding novel human membrane protein, useful for
PT identifying agonist, antagonist or modifiers or for producing
PT antibodies useful in therapeutic, diagnostic and pharmacogenomic
PT applications -
XX
XX Claim 4; Page 31; 32pp; English.
XX
XX The present sequence represents novel human membrane protein #4.
CC Human membrane protein #4 is 1 of 4 human membrane proteins
CC (AAU01210-AAU01213) given in the present invention. These membrane
CC proteins share structural similarity with membrane receptors such as
CC the IgE receptor and mammalian CD20. The novel human membrane proteins
CC are useful for identifying agonists, antagonists and modulators of the
CC membrane proteins. The membrane proteins can be used for diagnosis,
CC drug screening, pharmacogenomic applications, clinical trial monitoring
CC and the treatment of physiological disorders and diseases. The
CC polynucleotides encoding the membrane proteins can be used to generate
CC PCR primers or probes to identify mutations associated with a particular
XX disease.
XX
SQ Sequence 149 AA;

Query Match 71.6%; Score 735.5; DB 22; Length 149;
Best Local Similarity 74.5%; Pred. No. 1.7e-77;
Matches 149; Conservative 0; Mismatches 0; Indels 51; Gaps 1;

QY 1 MDSSTAHSPVFLVFPPEITASEYESTELSATTFSTQSPLOKLFARKMKILGTIQLFGIM 60
Db 1 mdsstahspvflvfppeitaseyestelsattfstqspqlkfarkmkilgtiqlfgim 60
QY 61 TFSFGVIFLTLKPYPRPFIFLSGYFPWGSVLFINSAGFLIAVKRKTETLLILSRIM 120
Db 61 tfsfgvifltlkpyprpfiflsgypfwgsvlfinsgafliavkrktet----- 112
QY 121 NFLSALGATAGIILLFGFTLDQNYICGYSHQNSQCKAVTVLFLGILITLMTFSIIEIFI 180
Db 113 -----lgilitlmtfsiieifi 129
QY 181 SLPFSILGCHSDCCECC 200
Db 130 slpfsilgchsdcccecc 149

RESULT 7
ABB12234
ID ABB12234 standard; peptide; 158 AA.
XX

CC membrane proteins. The membrane proteins can be used for diagnosis,
CC drug screening, pharmacogenomic applications, clinical trial monitoring
CC and the treatment of physiological disorders and diseases. The
CC polynucleotides encoding the membrane proteins can be used to generate
CC PCR primers or probes to identify mutations associated with a particular
CC disease.

XX
SQ Sequence 138 AA;

Query Match 56.7%; Score 582; DB 22; Length 138;
Best Local Similarity 92.1%; Pred. No. 1.le-59;
Matches 117; Conservative 4; Mismatches 0; Indels 6; Gaps 1;

QY 1 MDSSTAHPVFLVFPPEITASYESYESTELSGATTFSTQSPLOKLFARKMKILGTIQLFGIM 60
|||||
Db 1 mdsstahspvflvfppeitaseyestelsattfstqspklfarkmkilgtiqlfgim 60
|||||
QY 61 TSFSGVIFLFTLLKPYPRPFIFLSCYFPGSVFLFNSGAFLIAVKKRTTETLI----- 114
|||||
Db 61 tfsfgviflftllkpyprpfiflsgyfpwsvflfngsvflfngsagfliavkrkttetlikttfvv 120
|||||

QY 115 ILSRIMN 121
||:::
Db 121 iltkivs 127

RESULT 9
AAU01211
ID AAU01211 standard; Protein; 81 AA.

XX
AC AAU01211;

XX 26-SEP-2001 (first entry)

XX Novel human membrane protein #2.

XX Human; membrane protein; membrane receptor; IgE receptor; CD20;
XX physiological disorder.

XX Homo sapiens.

XX WO200146417-A2.

XX 28-JUN-2001.

XX 12-DEC-2000; 2000WO-US33742.

XX 22-DEC-1999; 99US-0171567.

XX (LEXI-) LEXICON GENETICS INC.

XX Walke DW, Turner CA;

XX WPI; 2001-408646/43.

XX N-PSDB; AAS04280.

XX Polynucleotide encoding novel human membrane protein, useful for
XX identifying agonist, antagonist or modifiers or for producing
XX antibodies useful in therapeutic, diagnostic and pharmacogenomic
XX applications -

XX Disclosure; Page 30; 32pp; English.

XX The present sequence represents novel human membrane protein #2.

XX Human membrane protein #2 is 1 of 4 human membrane proteins
XX (AAU01210-AAU01213) given in the present invention. These membrane
XX proteins share structural similarity with membrane receptors such as
XX the IgE receptor and mammalian CD20. The novel human membrane proteins
XX are useful for identifying agonists, antagonists and modulators of the
XX membrane proteins, and for producing antibodies specific to the
XX membrane proteins. The membrane proteins can be used for diagnosis,
XX drug screening, pharmacogenomic applications, clinical trial monitoring

CC and the treatment of physiological disorders and diseases. The
CC polynucleotides encoding the membrane proteins can be used to generate
CC PCR primers or probes to identify mutations associated with a particular
CC disease.

XX
SQ Sequence 81 AA;

Query Match 40.3%; Score 414; DB 22; Length 81;
Best Local Similarity 97.5%; Pred. No. 2e-40;
Matches 79; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 120 MNFSLAIGALAGIILLTFPGTLDQNYICGYSHONSQCKAVTVLFLGLITLMTFSLIELF 179
|||||
Db 1 mnlisalaraiagiilltfgildqnyicgyshnsgckavtvlfglilmtfsielf 60
|||||

QY 180 ISLFFSLGCHSEDCDCEQCC 200
|||||

Db 61 islpfslgchsedcdceqcc 81
|||||

RESULT 10

AAG03074

ID AAG03074 standard; Protein; 77 AA.

XX
AC AAG03074;

XX 06-OCT-2000 (first entry)

XX Human secreted protein, SEQ ID NO: 7155.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX N-PSDB; AAC03080.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX Claim 13; SEQ ID 7155; 71pp + CD-ROM; English.

XX The present sequence is a polypeptide encoded by one of a large number
XX of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
XX were prepared from total human RNAs or polyA+ RNAs derived from 30
XX different tissues. EST sequences usually correspond mainly to the 3'
XX untranslated region (UTR) of the mRNA because they are often obtained
XX from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
XX isolating cDNA sequences derived from the 5' ends of mRNAs and even in
XX those cases where longer cDNA sequences have been obtained, the full 5'
XX UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
XX ends and can therefore be used to obtain full length cDNAs and genomic
XX DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
XX chromosome mapping procedures. They are used to obtain upstream
XX regulatory sequences and to design expression and secretion vectors.

XX
SQ Sequence 77 AA;

Query Match 28.6%; Score 294; DB 21; Length 77;
Best Local Similarity 83.6%; Pred. No. 1.7e-26;
Matches 56; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

Qy	33	FTQSPLOQLPARKUKKILGTQILFGIMTFMSCGVFLTLLKPKYPRPPPFIFUSGYPFWGS	92
Db	11	fqkapekklfarkmkilgtiqxifgimxfstfgvifltlkkpyrpfifisxypfwgs	70
Qy	93	VLFINSG	99
Db	71	vlfinsg	77

RESULT 11
AAG89142
ID AAG89142 standard: Protein: 67 AA:

AC AAG89142;

DT 11-SEP-2001 (first entry)

Human secreted protein, SEQ ID NO: 262.

Human; secreted protein; gene therapy; vaccine; treatment; diagnosis;
KW
KW
KW GENSET.

OS Homo sapiens.

PN WO200142451-A2.

14-JUN-2001.

PF 07-DEC-2000; 2000WO-IB01938.

PR 08-DEC-1999; 99US-0169629.

XX
XX

XX
PI Dumas Milne Edwards J, Bougueleret L, Jobert S;

DR WPI; 2001-367870/38.
DR N-PSDB; AAH64745.

Full length GENSER human nucleic acids encoding potentially secreted proteins, useful in gene therapy and vaccination against a variety of diseases, and for diagnosis of those diseases -

PS Claim 21; Page 805; 921pp; English.

The invention relates to full length GENSET human nucleic acids encoding potentially secreted proteins. The nucleic acids and the polypeptides they encode may be used in the prevention, treatment and diagnosis of diseases associated with inappropriate GENSET gene expression. For example, they be used to treat disorders associated with decreased GENSET gene expression by rectifying mutations or deletions in a patient's genome that affect the activity of GENSET or by supplementing the patients own production of GENSET polypeptides. Conversely, antisense nucleic acid molecules may be administered to down regulate GENSET expression by binding with the cells' own genes and preventing their expression. The sense and antisense nucleic acids may also be used as DNA probes in diagnostic assays to detect and quantitate the presence of similar nucleic acid sequences in samples, and hence to determine which patients may be in need of restorative therapy. The GENSET polypeptides may be used as antigens in the production of antibodies and in assays to identify modulators (agonists and antagonists) of GENSET polypeptide expression and activity. The present sequence is a GENSET polypeptide of the invention.

Sequence	67 AA;
SQ	

Query Match 25.7%; Score 263.5; DB 22; Length 67;
Best Local Similarity 74.7%; Pred. No. 5.1e-23;
Matches 56; Conservative 3; Mismatches 7; Indels 9; Gaps 1;

```
Qy 1 mdsstahspvflvppeitasevesteiselsttftsqplqlfarkmkilgtqilfgim 60
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 mdsstahspvflvppeitasevesteiselsttftsqplqlfarkmkilgdh----- 54
```

```
QY      61 TFSFGVIFLFTLLKP 75
      | : | : |
db      55 ---sqalfcslilep 66
```

RESULT 12

AAW50743
ID AAW96745 standard; protein; 239 AA.

AA AAW96745;

DT 15-APR-1999 (first entry)

DE High affinity immunoglobulin E receptor-like protein (IGERB).

High affinity immunoglobulin E receptor-like protein ; IGERB;
 Incyte clone 927955; inflammatory response; AIDS; Addison's s disease;
 atherosclerosis; bronchitis; ulcerative colitis; diabetes mellitus;
 emphysema; gout; Graves's disease; osteoporosis; rheumatoid arthritis;
 Sjogren's syndrome; cancer.

OS Homo sapiens.

xx	Key	Location/Qualifiers
FH		

123	modified site	/note= "potential protein kinase C phosphorylation site"
124	FT	
125	FT	

FT	Modified site	IS7	FT	potential N-glycosylation site"
				/note=

192
PT
PT
PT

US5871930-A

16-FEB-1999.

21-AUG-1997: 97US-0916902.

AA 21-AUG-1997; 97US-0916902.

PA (INCY-) INCYTE PHARM INC.

PI Bandman O, Corley NC, Lal P;

DR WPI; 1999-166628/14.

XX

PT New polynucleotide coding for high affinity immunoglobulin E
PT receptor-like protein - and its complement, useful for diagnosis,
PT prevention and antisense therapy of inflammatory responses

PS Claim 1; Fig 1A-D; 30pp; English.

The present sequence represents a high affinity immunoglobulin E receptor-like protein (IGERB). The present sequence was first identified in Incyte clone 927955 for the brain cDNA library BRAIN004. Polynucleotides complementary to the IGERB cDNA can be used as probes to IGERB gene expression. IGERB proteins and nucleotides may be used for the diagnosis of inflammatory responses associated with expression of IGERB, e.g AIDS.

CC adenosine 5' disease, achrocytosis, mononucleus, ulcerative colitis
CC diabetes mellitus, emphysema, gout, Graves's disease, osteoporosis,
CC rheumatoid arthritis, Sjogren's syndrome and complications of cancer
CC The sequences may also be useful in assays that detect activation or

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 30, 2002, 00:43:57 ; Search time 49.1 Seconds
(without alignments)
3802.066 Million cell updates/sec

Title: US-09-821-821-1
Perfect score: 760
Sequence: 1 ttccagtcctccaggcagcc.....aaaaaaaaaaaaaaaaaaaaa 760

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	86.4	11.4	1669	2	US-08-916-902A-2
2	86.4	11.4	1669	2	US-09-213-389-2
3	61.6	8.1	1661	1	US-08-318-492-3
4	61.6	8.1	1661	1	US-08-707-340-3
5	61.6	8.1	1661	2	US-08-994-578-3
6	50.8	6.7	2545	1	US-07-869-933-22
7	50.8	6.7	2545	4	US-09-103-663-22
8	44.2	5.8	7218	1	US-08-232-463-14
9	42.4	5.6	1738	2	US-08-379-482A-2
10	41	5.4	658	4	US-09-385-982-327
11	40.4	5.3	3889	2	US-08-648-298-1
12	40.2	5.3	1897	1	US-08-184-632-1
13	39.8	5.2	247	1	US-08-691-641-4
14	39.2	5.2	2447	2	US-09-014-969-14
15	38.8	5.1	756	2	US-08-530-165-1
16	38.8	5.1	1172	1	US-07-945-288-9
17	38.8	5.1	1172	1	US-08-462-831-9
18	38.8	5.1	1172	1	US-08-461-809-9
19	38.8	5.1	1172	1	US-08-461-441-9
20	38.8	5.1	1172	5	PCT-US93-08518-9
21	38.4	5.1	991	3	US-08-924-747-25
22	38.4	5.1	991	4	US-09-247-373B-25
23	38.4	5.1	991	4	US-09-296-715-25
24	38.4	5.1	2028	2	US-09-211-930-12
25	38.4	5.1	2028	3	US-09-340-993-12
26	38.4	5.1	2028	4	US-09-468-442-12
27	38.4	5.1	84495	4	US-09-797-906-3

c

28	38.2	5.0	1359	4	US-09-387-574-11	Sequence 11, Appl
29	38.2	5.0	1359	4	US-09-668-096-11	Sequence 11, Appl
30	38	5.0	319	1	US-07-593-657-14	Sequence 14, Appl
31	38	5.0	587	1	US-08-313-681A-3	Sequence 3, Appl
32	38	5.0	587	3	US-09-322-911-3	Sequence 3, Appl
33	38	5.0	2381	1	US-08-021-608D-9	Sequence 9, Appl
34	38	5.0	2381	1	US-08-726-160-9	Sequence 9, Appl
35	38	5.0	2381	5	PCT-US94-01782-9	Sequence 9, Appl
36	38	5.0	2384	1	US-08-021-608D-1	Sequence 1, Appl
37	38	5.0	2384	1	US-08-726-160-1	Sequence 1, Appl
38	38	5.0	2384	5	PCT-US94-01782-1	Sequence 1, Appl
39	37.8	5.0	575	1	US-08-554-659-11	Sequence 11, Appl
40	37.4	4.9	1622	4	US-09-232-191-12	Sequence 12, Appl
41	37.4	4.9	1622	4	US-09-232-200-12	Sequence 12, Appl
42	37.4	4.9	1622	4	US-09-232-201-12	Sequence 12, Appl
43	37.4	4.9	1622	4	US-09-232-201-12	Sequence 12, Appl
44	37.4	4.9	2238	1	US-08-742-011-1	Sequence 1, Appl
45	37.4	4.9	2830	1	US-07-882-292-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-916-902A-2
; Sequence 2, Application US/08916902A
; Patent No. 5871930
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HIGH AFFINITY IMMUNOGLOBULIN E
; TITLE OF INVENTION: RECEPTOR-LIKE PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/916,902A
; FILING DATE: Herewith
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0371 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1669 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRAINOT04
; CLONE: 927955
; US-08-916-902A-2

```
Query Match      11.4%; Score 86.4; DB 2; Length 1669;
Best Local Similarity 51.1%; Pred. NO. 3.4e-13;
Matches 232; Conservative 0; Mismatches 216; Indels 6; Gaps 1;

Qy 216 aaaaaattatttgcagaaatgaaatcttagggactatccagatcctctgttggaaatta 275
Db 325 AGAAGTTCCTTGAAGGGAGAACCCAAAGTCTTGGGGTGTGCAGATTCTGACTGCCCTGA 384
Qy 276 tgaccttttcttggagttatcttcttcttcaactttgttaaaaccatatccagggttc 335
Db 385 TGAGCCCTTAGCATGGGAATAACAATGATGTGATGGCATCTAATACCTATGGAAGTAACC 444
Qy 336 cctttatatttccagatatactctgggctctgttttcttcaataattctggag 395
Db 445 CTATTTCCGTGTATATCGGGGTACACAAATTTGGGGTGTGATGTTATTTTCAGGAT 504
Qy 396 ccttcttaattgcagtgaaagaaacccagaaactctgataataattgagccgaataa 455
Db 505 CCTTGTCAATTCAGCAGGAAATAGAACTACAAAAGGCTGGTCCGAGGTAGTCTAGGAA 564
Qy 456 tgaattttcttagtgcctgggagcaatagctgaatcattctcctcacatttggttca 515
Db 565 TGAATATCACCAGCTCTGTACTGGCTGCATCAGGGATCTTAATCAACACATTTAGCTGG 624
Qy 516 tcttagat-----caaaactacatttgggttattctcaccacaaatagtcagtgtaag 569
Db 625 CGTTTTCATTCATCCATCCACCCCTTACTGTAACTACTATGGCAACTCAAAATAATTTGCATG 684
Qy 570 ctgttactgcctgttcttgggaatttgcattacattgatgacttccagcatttgaat 629
Db 685 GGACTATGTCATCTTAATGGGTCTGGATGGCATGGTCTCTTAAGTGTGCTGGAAT 744
Qy 630 tattcatttctgccttctcatttgggggtg 663
Db 745 TCTGCATGTGCTGCTCCTCTCTGCTGCTTGGATG 778

RESULT 2
US-09-213-389-2
; Sequence 2, Application US/09213389
; Patent No. 5977072
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HIGH AFFINITY IMMUNOGLOBULIN E
; TITLE OF INVENTION: RECEPTOR-LIKE PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/213,389
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/916,902
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0371 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555

; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1669 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRAINOT04
; CLONE: 927955
US-09-213-389-2

Query Match      11.4%; Score 86.4; DB 2; Length 1669;
Best Local Similarity 51.1%; Pred. NO. 3.4e-13;
Matches 232; Conservative 0; Mismatches 216; Indels 6; Gaps 1;

Qy 216 aaaaaattatttgcagaaatgaaatcttagggactatccagatcctctgttggaaatta 275
Db 325 AGAAGTTCCTTGAAGGGAGAACCCAAAGTCTTGGGGTGTGCAGATTCTGACTGCCCTGA 384
Qy 276 tgaccttttcttggagttatcttcttcttcaactttgttaaaaccatatccagggttc 335
Db 385 TGAGCCCTTAGCATGGGAATAACAATGATGTGATGGCATCTAATACCTATGGAAGTAACC 444
Qy 336 cctttatatttccagatatactctgggctctgttttcttcaataattctggag 395
Db 445 CTATTTCCGTGTATATCGGGGTACACAAATTTGGGGTGTGATGTTATTTTCAGGAT 504
Qy 396 ccttcttaattgcagtgaaagaaacccagaaactctgataataattgagccgaataa 455
Db 505 CCTTGTCAATTCAGCAGGAAATAGAACTACAAAAGGCTGGTCCGAGGTAGTCTAGGAA 564
Qy 456 tgaattttcttagtgcctgggagcaatagctgaatcattctcctcacatttggttca 515
Db 565 TGAATATCACCAGCTCTGTACTGGCTGCATCAGGGATCTTAATCAACACATTTAGCTGG 624
Qy 516 tcttagat-----caaaactacatttgggttattctcaccacaaatagtcagtgtaag 569
Db 625 CGTTTTCATTCATCCATCCACCCCTTACTGTAACTACTATGGCAACTCAAAATAATTTGCATG 684
Qy 570 ctgttactgcctgttcttgggaatttgcattacattgatgacttccagcatttgaat 629
Db 685 GGACTATGTCATCTTAATGGGTCTGGATGGCATGGTCTCTTAAGTGTGCTGGAAT 744
Qy 630 tattcatttctgccttctcatttgggggtg 663
Db 745 TCTGCATGTGCTGCTCCTCTCTGCTGCTTGGATG 778

RESULT 3
US-08-318-492-3
; Sequence 3, Application US/08318492
; Patent No. 5552312
; GENERAL INFORMATION:
; APPLICANT: Lim, Bing
; APPLICANT: Adra, Chaker N.
; APPLICANT: Lelias, Jean-Michel
; TITLE OF INVENTION: RECOMBINANT HTM4 GENE, PROTEIN AND
; TITLE OF INVENTION: ASSAYS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/318,492
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Brook, David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: BIH94-03
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1661 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 97..741
; US-08-318-492-3

Query Match      8.1%; Score 61.6; DB 1; Length 1661;
Best Local Similarity 50.6%; Pred. No. 7.1e-07;
Matches 238; Conservative 0; Mismatches 214; Indels 18; Gaps 3;

QY 232 aaaaatgaaatcttagggactccagatcctgtgttggaattatgacctttctcttgg 291
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 234 AAAATTACAAGTCTTGGGGCCATCCAGATCCTGAATGCAGCAATGATTCGGCTTGGG 293
QY 292 agt-----tatctctcttctcaacttggtaaaacacatccaaggttcccttat 342
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 294 TGCTTTCTGGGTTCCTTGGCAATACCCATACCCTTCCAAAAGCACTTCCTTTCTTCAC 353
QY 343 attcttcaggatataccattctgtgggctctgtttgttccattaaattctggagccttct 402
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 354 CTCTACACAGGCTACCCGATTTGGGGTGTGTGTTTTCTGTAGTTCAGGAACCTTGC 413
QY 403 aattgcagtgaaagaaacacacagaaactctgataattatgagccgaataatgaattt 462
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 414 TGTTGTAGCAGGGATAAACCCCAAGAACATGGATACAGAACAGTTCCTTGGAAATGAACAT 473
QY 463 tcttagtgcctgggagcaatagctggaatcattctctccacatttgggttccatccataga 522
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 474 TGCCAGTGTACAAATTCGACTAGTGGGACTGCTTTCTCTACTAAATATAGCAGTTAA 533
QY 523 tcaaaactacatttgggttat-tctcaccacaaataatgtagtgaaggtgttact---- 577
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 534 TATCCAGTCATTAAAGGAGTTGTACCTTCATCAGAGTCACCGGACCTATGCAATTACAT 593
QY 578 -----gtcctgttcttgggaattttgattacattgacgttccagcattattgaattatt 633
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 594 GGGCTCCATATCAAAATGGCATGGTGTCTACTGCTGATCTCACCTTGTGGAATTTATG 653
QY 634 cattctctgcctttctcaattttgggggtgcactcagagattgtgatt 683
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 654 CGTAACATATCTACCATAGCCATGTGTGCAATGCCAACCTGTCTGTAATT 703

RESULT 4
US-08-707-340-3
; Sequence 3, Application US/08/07340
; Patent No. 5705615
; GENERAL INFORMATION:
; APPLICANT: Lim, Bing
; APPLICANT: Adra, Chaker N.
; APPLICANT: Lelias, Jean-Michel
; TITLE OF INVENTION: RECOMBINANT HTM4 GENE, PROTEIN AND
; TITLE OF INVENTION: ASSAYS
; NUMBER OF SEQUENCES: 4
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/707,340
; FILING DATE: 03-SEP-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/318,492
; FILING DATE: 06-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/675,648
; FILING DATE: 03-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Brook, David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: BIH94-03A2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1661 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 97..738
; US-08-707-340-3

Query Match      8.1%; Score 61.6; DB 1; Length 1661;
Best Local Similarity 50.6%; Pred. No. 7.1e-07;
Matches 238; Conservative 0; Mismatches 214; Indels 18; Gaps 3;

QY 232 aaaaatgaaatcttagggactccagatcctgtgttggaattatgacctttctcttgg 291
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 234 AAAATTACAAGTCTTGGGGCCATCCAGATCCTGAATGCAGCAATGATTCGGCTTGGG 293
QY 292 agt-----tatctctcttctcaacttggtaaaacacatccaaggttcccttat 342
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 294 TGCTTTCTGGGTTCCTTGGCAATACCCATACCCTTCCAAAAGCACTTCCTTTCTTCAC 353
QY 343 attcttcaggatataccattctgtgggctctgtttgttccattaaattctggagccttct 402
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 354 CTCTACACAGGCTACCCGATTTGGGGTGTGTGTTTTCTGTAGTTCAGGAACCTTGC 413
QY 403 aattgcagtgaaagaaacacacagaaactctgataattatgagccgaataatgaattt 462
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 414 TGTTGTAGCAGGGATAAACCCCAAGAACATGGATACAGAACAGTTCCTTGGAAATGAACAT 473
QY 463 tcttagtgcctgggagcaatagctggaatcattctctccacatttgggttccatccataga 522
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 474 TGCCAGTGTACAAATTCGACTAGTGGGACTGCTTTCTCTACTAAATATAGCAGTTAA 533
QY 523 tcaaaactacatttgggttat-tctcaccacaaataatgtagtgaaggtgttact---- 577
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 534 TATCCAGTCATTAAAGGAGTTGTACCTTCATCAGAGTCACCGGACCTATGCAATTACAT 593
QY 578 -----gtcctgttcttgggaattttgattacattgacgttccagcattattgaattatt 633
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 594 GGGCTCCATATCAAAATGGCATGGTGTCTACTGCTGATCTCACCTTGTGGAATTTATG 653
QY 634 cattctctgcctttctcaattttgggggtgcactcagagattgtgatt 683
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 654 CGTAACATATCTACCATAGCCATGTGTGCAATGCCAACCTGTCTGTAATT 703
```


Query Match	5.3%;	Score 40.2;	DB 1;	Length 1897;
Best Local Similarity	60.6%;			
Pred. NO.	0.21;			
Mismatches	0.	Mismatches	43:	Indels 0:
Matches	56	Conservative		Gaps 0:

QY	652	aattttggsgtgcactcagagatgttgattgtgaaacaattgttgcactagcaactgtg	711
Db	1767	AGTTTGGGACACAGTTGATTATTATGTTTATACAAATTTTAAATAATGAATAGGT	1826
QY	712	agaataaagatgtgttaaactcaaaaaaataaaaaaataaaaaa 760	
Db	1827	AGTATCATATATATGAAAAAATAAAAAAATAAAAAAATAAAAAA 1875	

```

RESULT 13
US-08-691-641-4
; Sequence 4, Application US/08691641
; Patent No. 5728819
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John
; APPLICANT: Lavallee, Edward
; APPLICANT: Racie, Lisa
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Evans, Cheryl
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING THEM
; NUMBER OF INVENTIONS: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
;

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/691,641
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 247 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cdna
US-08-691-641-4

Query Match 5.2%; Score 39.8; DB 1; Length 247;
Best Local Similarity 65.9%; Pred. No. 0.13;
Matches 56; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

Qy 676 ttgtgattgtgaacaatgttgttgactgacactgtgagaataaagatgtgttaaatctc 735
Db 163 TTAGGAGTGTATTTTCTACTATGTGTGGCAGGATACATAAAGANTTTTACAAANTA 222

Qy 736 aaaaaaaaaaaaaaaaaaaaaa 760
Db 223 AAAAAAAAAAAAAAAAAAAAAA 247

RESULT 14

US-09-014-969-14
Sequence 14, Application US/09014969
Patent No. 5965397
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John M.
APPLICANT: Lavallie, Edward R.
APPLICANT: Racie, Lisa A.
APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Spaulding, Vikki
APPLICANT: Agostino, Michael J.
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/014,969
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Sprunger, Suzanne A.
REGISTRATION NUMBER: 41,323
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8284
TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 2447 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cdna
US-09-014-969-14

Query Match 5.2%; Score 39.2; DB 2; Length 2447;
Best Local Similarity 51.1%; Pred. No. 0.4;
Matches 89; Conservative 1; Mismatches 84; Indels 0; Gaps 0;

Qy 587 ttgggaattttgattacattgatcttcacgatttgaattattcattctctgcct 646
Db 2118 TTAGAATGTATAGATAGGTTTGTGAAAGTCTAAATAATGGCTGTATAGTATATATAT 2177
Qy 647 ttctcaattttgggtgccactcagagagattgtgtggaacaattgtgtgactagca 706
Db 2178 GGTTCACATATCTGGATCTGTGATTTTGATTTTGTACTTTTAATGTGACAAATAACCTT 2237
Qy 707 ctgtgagaataaagatgtgttaaatctcaaaaaaaaaaaaaaaaaaaaaa 760
Db 2238 TTGGGAGAAA 2291

RESULT 15

US-08-530-165-1
Sequence 1, Application US/08530165
Patent No. 5907081
GENERAL INFORMATION:
APPLICANT: Isaac, Peter G.
APPLICANT: Roberts, Jeremy A.
APPLICANT: Coupe, Simon A.
TITLE OF INVENTION: Control of Plant Abscission and Pod Dehiscence
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 New York Avenue, N.W., Suite 600
CITY: Washington
STATE: DC
COUNTRY: 20005
ZIP: USA
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/530,165
FILING DATE: 29-SEP-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB94/00689
FILING DATE: 31-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9306726.2
FILING DATE: 31-MAR-1993
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 756 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
FEATURE:
NAME/KEY: CDS
LOCATION: 15..458
US-08-530-165-1

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 30, 2002, 01:33:42 ; Search time 214.01 Seconds
(without alignments)
6097.168 Million cell updates/sec

Title: US-09-821-821-1

Perfect score: 760

Sequence: 1 ttccagtgctccagcgcc.....aaaaaaaaaaaaaaaaaaaaa 760

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_032802.*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	760	100.0	760	22	Human CD20/IgE-rec
2	730.2	96.1	747	22	Human secreted pro
3	689.8	90.8	737	22	Human reproductive
4	668.4	87.9	689	22	Human immune syste
5	666.4	87.7	676	22	DNA sequence encod
6	635.8	83.7	695	22	Human secreted pro
7	598.2	78.7	603	22	Novel human membra
8	351.2	46.2	372	22	Human secreted pro
9	347.4	45.7	450	22	Novel human membra

10	340.2	44.8	417	22	AAS04281	Novel human membra
11	242.8	31.9	246	22	AAS04280	Novel human membra
12	233.4	30.7	468	22	AAH64745	Human secreted pro
13	86.4	11.4	1330	21	AAAI6693	Human secreted pro
14	86.4	11.4	1669	20	AAI5104	High affinity immu
15	86.4	11.4	1670	21	AAZ32842	Human high affinity
16	84.8	11.2	1677	22	AAK94442	Human full-length
17	78.8	10.4	822	22	AAK92131	Human cDNA 5'-end
18	78.8	10.4	822	22	AAK93395	Human cDNA clone r
19	69.4	9.1	562	22	ABA09445	Human IGERB homolo
20	64.4	8.5	1710	23	AAK83423	DNA encoding novel
21	62.6	8.2	1563	22	AAZ36238	CDNA encoding a bo
22	61.8	8.1	708	22	AAF77694	Murine wild-type F
23	61.6	8.1	1661	17	AAI45120	HTM4 protein codin
24	61.6	8.1	1661	19	AAV03875	HTM4 gene. Homo s
25	59.8	7.9	1474	13	AAO21170	Clone CD20.4 encod
26	58.2	7.7	1473	10	AAK90610	CD20.4 cDNA. AAN
27	58.2	7.7	1474	17	AAI4710	Human CD20.4 anti
28	58.2	7.7	1474	19	AAV63448	Human CD20.4 anti
29	58.2	7.7	1474	20	AAV81206	Human CD20.4 anti
30	58.2	7.7	1474	21	AAAS0584	Human cell surface
31	58.2	7.7	1474	22	AAO31178	Human lymphocyte c
32	58.2	7.7	1476	10	AAK90613	CD20 cDNA. AAN90
33	58.2	7.7	1476	13	AAO21173	Human CD20 antigen
34	58.2	7.7	1476	17	AAI4713	Human CD20 antigen
35	58.2	7.7	1476	19	AAV63451	Human CD20 antigen
36	58.2	7.7	1476	20	AAV81209	Human CD20 antigen
37	58.2	7.7	1476	21	AAAS0587	Human cell surface
38	58.2	7.7	1476	22	AAO31181	Human lymphocyte c
39	54.6	7.2	1060	20	AAK97730	Extended human sec
40	54.2	7.1	932	22	AAI4723	Dendritic cell (DC
41	54.2	7.1	945	19	AAV59668	Human secreted pro
42	54.2	7.1	1001	21	AAZ52580	Human secreted pro
43	54.2	7.1	1036	22	AAI4722	Dendritic cell (DC
44	54.2	7.1	1106	22	ABA09233	Human secreted pro
45	54.2	7.1	1106	22	AAK93750	Human protein enco

ALIGNMENTS

RESULT 1

AAI21441
ID AAD21441 standard; cDNA; 760 BP.

XX AAD21441;

XX 28-JAN-2002 (first entry)

XX Human CD20/IgE-receptor like protein encoding cDNA, agp-96614-al.

XX Human; CD20/IgE-receptor like protein; immunoglobulin E; agp-96614-al;
KW agp-69406-al; cancer; abnormal cell proliferation; autoimmune disease;
KW ovarian cancer; brain cancer; arteriosclerosis; vascular stenosis;
KW rheumatoid arthritis; multiple sclerosis; allergy; dermatitis; asthma;
KW reproductive disease; diabetes; transplant rejection; endometriosis;
KW infertility; gene therapy; ss.

XX Homo sapiens.

XX Key Location/Qualifiers
XX CDS 98..700

XX FT /tag= a
XX FT /product= "Human CD20/IgE-receptor like protein,
XX FT agp-96614-al"

XX W0200174903-A2.

XX 11-OCT-2001.

XX 29-MAR-2001; 2001WO-US10048.

XX 30-MAR-2000; 2000US-193728P.

PR 27-NOV-2000; 2000US-0723258.
PA (AMGE-) AMGEN INC.
PI Welcher AA, Calzone FJ;
XX
DR WPI: 2001-662968/76.
DR P-PSDB; AAE13062.
XX
PT Novel CD20/IgE-receptor like polypeptides and polynucleotides,
PT antagonists and antibodies of the polypeptide useful for treating
PT ameliorating or preventing diseases associated with the polypeptide
PT e.g. cancer, asthma
XX
PS Claim 1; Fig 1; 145pp; English.
XX
CC The invention relates to human CD20/immunoglobulin E (IgE)-receptor
CC like polypeptides designated as agp-96614-al and agp-69406-al and
CC nucleic acid molecules encoding such polypeptides. Polypeptides of
CC the invention are useful for treating, preventing or ameliorating
CC a disease, condition, or disorder which includes cancer such as
CC brain cancer, ovarian cancer; abnormal cell proliferation such as
CC arteriosclerosis, vascular restenosis; pathology from allergens
CC such as allergies, asthma, dermatitis; dysfunction of immune system
CC such as rheumatoid arthritis, autoimmune disease, multiple sclerosis,
CC diabetes, transplant rejection and reproductive diseases such as
CC infertility, preterm labour and delivery, endometriosis etc. They
CC are also useful for identifying antagonists and as immunogens, for
CC raising antibodies which may also be used to prevent, treat or
CC diagnose a number of diseases and disorders. Polynucleotides of the
CC invention are used to map the location of CD20/IgE-receptor like
CC gene and related genes on chromosomes and as hybridisation probes.
CC They are also useful in gene therapy. The present sequence is
CC human CD20/IgE-receptor like protein encoding cDNA, agp-96614-al.
XX
SQ Sequence 760 BP; 232 A; 157 C; 126 G; 245 T; 0 other;

Query Match 100.0%; Score 760; DB 22; Length 760;
Best Local Similarity 100.0%; Pred. No. 5.2e-165;
Matches 760; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ttccagtgctccaggcagctccagcacagaagaacatggtctagactgaagtaccaa 60
Db 1 ttccagtgctccaggcagctccagcacagaagaacatggtctagactgaagtaccaa 60
Qy 61 cttaaatcatctcttccaaattatccaccacacatcatggtattcgaagccgcacacag 120
Db 61 cttaaatcatctcttccaaattatccaccacacatcatggtattcgaagccgcacacag 120
Qy 121 tccgggtgttctgtatttctccagaaatcaactgcttcagaatatgagtcacagaact 180
Db 121 tccgggtgttctgtatttctccagaaatcaactgcttcagaatatgagtcacagaact 180
Qy 181 ttccagcacagcttttcaactcaaacgcttgcacaaataattattgctagaataatgaa 240
Db 181 ttccagcacagcttttcaactcaaacgcttgcacaaataattattgctagaataatgaa 240
Qy 241 aattctggagactatccagatcctgtttggaattatgacctttcttttggagtattctt 300
Db 241 aattctggagactatccagatcctgtttggaattatgacctttcttttggagtattctt 300
Qy 301 ccttttcaatttgaataacatccagaggttccctttatatttcttccagagatacc 360
Db 301 ccttttcaatttgaataacatccagaggttccctttatatttcttccagagatacc 360
Qy 361 attctggggctgtttgttctattattctggagccttccactaatgctgaaagaaga 420
Db 361 attctggggctgtttgttctattattctggagccttccactaatgctgaaagaaga 420
Qy 421 aaccacagaactctgataatattgagccgaaataatgaattttcttagtgccttggagc 480
Db 421 aaccacagaactctgataatattgagccgaaataatgaattttcttagtgccttggagc 480

QY 481 aatagctgggaatcattctcctcacatttgggtttcattcctcctagatcaaaactacatttgg 540
Db 481 aatagctgggaatcattctcctcacatttgggtttcattcctcctagatcaaaactacatttgg 540
QY 541 ttattctcaccacaaatagtcagtgtaaggctgttactgtctcttcttcttgggaatttggat 600
Db 541 ttattctcaccacaaatagtcagtgtaaggctgttactgtctcttcttcttgggaatttggat 600
QY 601 tacatgtagactttcagcaattattgaattattcatttcttctccttctcatttgg 660
Db 601 tacatgtagactttcagcaattattgaattattcatttcttctccttctcatttgg 660
QY 661 gtgccactcagagattgtgattgtgaacaattgtttgttactgactgactgtgagaataag 720
Db 661 gtgccactcagagattgtgattgtgaacaattgtttgttactgactgactgtgagaataag 720
QY 721 atgtgttaaaatctcaaaaaaataaaaaaataaaaaaataaaaaa 760
Db 721 atgtgttaaaatctcaaaaaaataaaaaaataaaaaaataaaaaa 760
RESULT 2
AAH64741
ID AAH64741 standard; cDNA; 747 BP.
XX
AC AAH64741;
XX
DT 11-SEP-2001 (first entry)
XX
DE Human secreted protein cDNA, SEQ ID NO: 17.
XX
KW Human; secreted protein; gene therapy; vaccine; treatment; diagnosis;
KW GENSET; ss.
XX
OS Homo sapiens.
XX
PN WO200142451-A2.
XX
PD 14-JUN-2001.
XX
PF 07-DEC-2000; 2000WO-IB01938.
XX
PR 08-DEC-1999; 990US-0169629.
PR 06-MAR-2000; 2000US-0187470.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Bougueleret L, Jobert S;
XX
DR WPI: 2001-367870/38.
DR P-PSDB; AAG89138.
XX
PT Full length GENSET human nucleic acids encoding potentially secreted
PT proteins, useful in gene therapy and vaccination against a variety of
PT diseases, and for diagnosis of those diseases -
XX
PS Claim 7; Page 576-577; 921pp; English.
XX
CC The invention relates to full length GENSET human nucleic acids encoding
CC potentially secreted proteins. The nucleic acids and the polypeptides
CC they encode may be used in the prevention, treatment and diagnosis of
CC diseases associated with inappropriate GENSET gene expression. For
CC example, they be used to treat disorders associated with decreased
CC GENSET gene expression by rectifying mutations or deletions in a
CC patient's genome that affect the activity of GENSET or by supplementing
CC the patients own production of GENSET polypeptides. Conversely,
CC antisense nucleic acid molecules may be administered to down regulate
CC GENSET expression by binding with the cells' own genes and preventing
CC their expression. The sense and antisense nucleic acids may also be
CC used as DNA probes in diagnostic assays to detect and quantitate the
CC presence of similar nucleic acid sequences in samples, and hence to
CC determine which patients may be in need of restorative therapy.

The GENSET polypeptides may be used as antigens in the production of antibodies and in assays to identify modulators (agonists and antagonists) of GENSET polypeptide expression and activity. The present sequence is a GENSET nucleic acid of the invention.

QY 707 ctgtgagaataagatgtgttaaaatctcaaaaaaataaaaaaataaaaaa 759
Db 78 CTGTGAGATAAGATGTGTAAATATTATGTAAATAAAAAAATAAAAAA 26

RESULT 4
AAD18275
ID AAD18275 standard; cDNA; 689 BP.
XX AAD18275;
AC
XX
DT 18-DEC-2001 (first entry)
DE Human immune system-related protein-encoding gene 9 cDNA clone HTEN45.
XX Human; immune system-related protein; allergy; rheumatoid arthritis;
KW cancer; Gaucher's disease; viral hepatitis; gene therapy; cystostatic;
KW diabetes mellitus; arrhythmia; wound healing; ischaemic lesion; AIDS;
KW Acquired Immune Deficiency Syndrome; viricide; hepatotropic; vasotropic;
KW autoimmune disorder; inflammation; cardiovascular disorder; hair loss;
KW wound healing; cell proliferation; skin aging; endocrine disorder;
KW food preservative; ss.
XX Homo sapiens.
XX
XX
FH Key Location/Qualifiers
FT CDS 29..634
FT /tag=a
FT /product="Human immune-system related protein"
XX WO200166722-A1.
XX 13-SEP-2001.
XX
XX 07-MAR-2001; 2001WO-US07260.
XX
XX 08-MAR-2000; 2000US-187873P.
XX 11-AUG-2000; 2000US-224367P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX NI J, Hilbert D, Kenny JJ, Moore PA, Choi GH, Soppet DR, Ebner R;
XX Gruber JR, Endress GA, Ruben SM;
XX
XX WPT: 2001-589939/66.
XX P-PSDB: AAE10917.
XX
XX Novel isolated immune system-related polypeptide useful for treating
XX rheumatoid arthritis, AIDS, allergy, cancer, Gaucher's disease,
XX diabetes mellitus, arrhythmia, wound healing, ischemic lesions and
XX viral hepatitis
XX
XX Claim 1; Page 301; 315pp; English.
XX
XX The invention relates to human immune system-related protein and their
XX DNA. Human immune-system related protein and DNA are useful for
XX preventing, treating or ameliorating a medical condition in a mammalian
XX subject, for diagnosing, preventing or treating immune system-associated
XX disorders, autoimmune disorders (rheumatoid arthritis), inflammatory
XX disorders (allergies), immunodeficiencies (AIDS), inflammatory disorders
XX (allergies), infectious diseases (e.g., viral hepatitis), complement
XX activation disorders, immune complex diseases, neoplastic disorders
XX (cancer), hyperproliferative disorders (Gaucher's disease), disorders
XX associated with neovascularisation, diseases at the cellular level,
XX cardiovascular disorders (arrhythmias), wound healing and epithelial
XX cell proliferation, endocrine disorders (diabetes mellitus) and
XX neurological disorders (ischaemic lesions). Immune-system related protein
XX or DNA is useful for preventing hair loss, skin aging due to sunburn, to
XX maintain organs before transplantation, to treat weight disorders, to
XX modulate mammalian characteristics, to change a mammal's mental or
XX physical state, or as a food additive or preservative. Immune-system
XX related DNA is useful in gene therapy, for chromosome identification.
XX radiation hybrid mapping, long range restriction mapping and in forensic

CC bioloq. The present sequence represents a human immune-system related
CC protein-encoding cDNA of the invention.
XX
SQ Sequence 689 BP; 204 A; 138 C; 114 G; 233 T; 0 other;

Query Match 87.9%; Score 668.4; DB 22; Length 689;
Best Local Similarity 99.0%; Pred. No. 5.1e-144;
Matches 683; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

QY 70 ctctttcaaatattatcacccagacacacatcatgattcaagcagccagcacagctcgggtgtt 129
Db 1 ctctttcaaatattatcacccagacacacatcatgattcaagcagccagcacagctcgggtgtt 60
QY 130 tctgtatttctccagaaatcaactgcttcagaatatgattccagacagaaatcttcagccac 189
Db 61 tctgtatttctccagaaatcaactgcttcagaatatgattccagacagaaatcttcagccac 120
QY 190 gacctttcaactcaaaagcccttcgcaaaaattatttctgtagaaaaatgaaaaatcttagg 249
Db 121 gacctttcaactcaaaagcccttcgcaaaaattatttctgtagaaaaatgaaaaatcttagg 180
QY 250 gactatccagatcctgtttggaattatgacctttcttttggagtattcttcttttccac 309
Db 181 gactatccagatcctgtttggaattatgacctttcttttggagtattcttcttttccac 240
QY 310 ttgttaaaaccatataccaaggtttcccttttatatttttcagagatatccattctgggg 369
Db 241 ctgttaaaaccatataccaaggtttcccttttatatttttcagagatatccattctgggg 300
QY 370 ctctgtttgttctaatattctggagccttctctaatattcagtgaaaaaagaaacacaga 429
Db 301 ctctgtttgttctaatattctggagccttctctaatattcagtgaaaaaagaaacacaga 360
QY 430 aactctgataattatgagccgaataatgaattttcttctgagccctggagcaatagctgg 489
Db 361 aactctgataattatgagccgaataatgaattttcttctgagccctggagcaatagctgg 420
QY 490 aatcattctctcaacatttgggtttctctatctatagatcaaaactacatttgggtttattctca 549
Db 421 aatcattctctcaacatttgggtttctctatctatagatcaaaactacatttgggtttattctca 480
QY 550 caaaatagctagtgtaaggctgttactgtctgttcttcttggaattttgattacattgat 609
Db 481 ccaaaatagctagtgtaaggctgttactgtctgttcttcttggaattttgattacattgat 540
QY 610 gactttcagcattattgaattattctctgtcttctcaatttttgggtggcactc 669
Db 541 gactttcagcattattgaattattctctgtcttctcaatttttgggtggcactc 599
QY 670 agaggattgtgattgtgaacaattgttggtagcactgtgagaataaagatgtgttaa 729
Db 600 agaggattgtgattgtgaacaattgttggtagcactgtgagaataaagatgtgttaa 659
QY 730 aatctcaaaaaaataaaaaaataaaaaa 759
Db 660 aatattatgtaaaaaaataaaaaaataaaaaa 689

RESULT 5
AAS04283
ID AAS04283 standard; DNA; 676 BP.
XX
XX AAS04283;
AC
XX
DT 26-SEP-2001 (first entry)
XX
XX DNA sequence encoding novel human membrane protein.
DE Human; membrane protein; membrane receptor; IgE receptor; CD20;
KW physiological disorder; ds.
XX
XX Homo sapiens.
OS

XX WO200146417-A2.
XX 28-JUN-2001.
XX 12-DEC-2000; 2000WO-US33742.
XX 22-DEC-1999; 99US-0171567.
XX (LEXI-) LEXICON GENETICS INC.
XX Walke DW, Turner CA;
XX WPI; 2001-408646/43.
XX Polynucleotide encoding novel human membrane protein, useful for
PT identifying agonist, antagonist or modifiers or for producing
PT antibodies useful in therapeutic, diagnostic and pharmacogenomic
PT applications -
XX Disclosure; Page 31-32; 32pp; English.
XX The present sequence encoding for a novel human membrane protein
CC includes the flanking 5'- and 3'- sequences. Four amino acid sequences
CC for novel human membrane proteins (AAU01210-AAU01213) are given in the
CC present invention. These membrane proteins share structural similarity
CC with membrane receptors such as the Ige receptor and mammalian CD20.
CC The novel human membrane proteins are useful for identifying agonist,
CC antagonist and modulators of the membrane proteins, and for producing
CC antibodies specific to the membrane proteins. The membrane proteins can
CC be used for diagnosis, drug screening, pharmacogenomic applications,
CC clinical trial monitoring and the treatment of physiological disorders
CC and diseases. The polynucleotides encoding the membrane proteins can be
CC used to generate PCR primers or probes to identify mutations associated
XX with a particular disease.
SQ Sequence 676 BP; 193 A; 140 C; 112 G; 231 T; 0 other;

Query Match 87.7%; Score 666.4; DB 22; Length 676;
Best Local Similarity 99.1%; Pred. No. 1.5e-143;
Matches 670; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 66 tcattctcttcaaatatcacccagacacccatcatgattcgaagcaccgacacagtcgg 125
Db 1 tcattctcttcaaatatcacccagacacccatcatgattcgaagcaccgacacagtcgg 60
Qy 126 tgtttctggtatttctccagaaatcactgtcttcagaatagatgagtcacagaaactttcag 185
Db 61 tgtttctggtatttctccagaaatcactgtcttcagaatagatgagtcacagaaactttcag 120
Qy 186 ccacgaacttttcaactcaagcccttgcaaaaattattgttagaaaaatgaaaatct 245
Db 121 ccacgaacttttcaactcaagcccttgcaaaaattattgttagaaaaatgaaaatct 180
Qy 246 tagggactatccagatcctgtttggaattatgacctttcttttgagttatcttcttt 305
Db 181 tagggactatccagatcctgtttggaattatgacctttcttttgagttatcttcttt 240
Qy 306 tcacttgttaaaaccatataccaggtttccctttatatttctttcaggatattcattct 365
Db 241 tcacttgttaaaaccatataccaggtttccctttatatttctttcaggatattcattct 300
Qy 366 ggggctgtttgttcaataatctggagccttcttaattgagtgaaaagaaaaacca 425
Db 301 ggggctgtttgttcaataatctggagccttcttaattgagtgaaaagaaaaacca 360
Qy 426 cagaaactctgataatattgagccggaataatgattttcttagtgcctgggagcaatag 485
Db 361 cagaaactctgataatattgagccggaataatgattttcttagtgcctggagcaatag 420
Qy 486 ctggaatcatctcctcaatttgggttctcatctcagatacaaaactacatttgggtatt 545

Db 421 ctggaatcattctcctcacatttgggtttcattcatcctagatcaaaactacatttgggtatt 480
Qy 546 ctacacaaatagtcagtgtaagctgttactgtctcttcttcttgggaatttggattacat 605
Db 481 ctacacaaatagtcagtgtaagctgttactgtctcttcttcttgggaatttggattacat 540
Qy 606 tgatgactttcagcattattgaattattcattcttctgcttctcatttctcatttgggtgccc 665
Db 541 tgatgactttcagcattattgaattattcattcttctgcttctcatttctcatttgggtgccc 600
Qy 666 actcagagattgtgattgtgaacaattgttggactagcactgtgagataaagatgtg 725
Db 601 actcagagattgtgattgtgaacaattgttggactagcactgtgagataaagatgtg 660
Qy 726 ttaaaatctcaaaaaa 741
Db 661 ttaaaataaaaaaaa 676

RESULT 6
ABA09478
ID ABA09478 standard; cDNA; 695 BP.
XX
AC ABA09478;
XX
DT 11-JAN-2002 (first entry)
XX
DE Human secreted protein homologue-encoding cDNA, SEQ ID NO:1254.
XX
KW Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; antiinflammatory;
KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW cytostatic; osteopathic; vasotrophic; cardiant; virucide; antibacterial;
KW antifungal; vulnery; antiulcer; ss.
XX
XX Homo sapiens.
OS
XX WO200157188-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 05-FEB-2001; 2001WO-US03800.
PF
XX
XX 03-FEB-2000; 2000US-0496914.
PR
XX 27-APR-2000; 2000US-0560875.
PR
XX (HYSE-) HYSEQ INC.
PA
XX
XX Tang Yt, Liu C, Drmanac RT;
PI
XX WPI; 2001-457740/49.
DR
XX P-PSDB; ABB12234.
PI
XX Human proteins and DNA encoding sequences useful for preventing,
PT treating or ameliorating a medical condition in a mammalian subject
PT e.g. arthritis and cancer -
XX
XX Claim 1; Page 961; 1963pp; English.
XX
XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC invention also relates to vectors and recombinant host cells comprising a
CC nucleotide of the invention, methods of producing the novel polypeptides,
CC antibodies against the polypeptides, methods of detecting the nucleotides
CC or polypeptides in a sample, and methods of identifying compounds which

XX The invention relates to full length GENSET human nucleic acids encoding
CC potentially secreted proteins. The nucleic acids and the polypeptides
CC they encode may be used in the prevention, treatment and diagnosis of
CC diseases associated with inappropriate GENSET gene expression. For
CC example, they be used to treat disorders associated with decreased
CC GENSET gene expression by rectifying mutations or deletions in a
CC patient's genome that affect the activity of GENSET or by supplementing
CC the patients' own production of GENSET polypeptides. Conversely,
CC antisense nucleic acid molecules may be administered to down regulate
CC GENSET expression by binding with the cells' own genes and preventing
CC their expression. The sense and antisense nucleic acids may also be
CC used as DNA probes in diagnostic assays to detect and quantitate the
CC presence of similar nucleic acid sequences in samples, and hence to
CC determine which patients may be in need of restorative therapy.
CC The GENSET polypeptides may be used as antigens in the production of
CC antibodies and in assays to identify modulators (agonists and
CC antagonists) of GENSET polypeptide expression and activity. The
CC present sequence is a GENSET nucleic acid of the invention.

XX Sequence 468 BP: 155 A: 94 C: 81 G: 138 T: 0 other:
XX

[illegible]

RESULT	13
AAA16693	
ID	AAA16693 standard; cDNA; 1330 BP.
XX	
AC	AAA16693;
XX	
DT	16-JUN-2000 (first entry)

XX Human; secreted protein; immunestimulant; immunosuppressant; virucide;
 KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
 KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
 KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
 KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
 KW connective tissue disease; multiple sclerosis; erythematosis;
 KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
 KW Gullain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
 KW insulin dependent diabetes mellitus; graft-versus-host-disease;
 KW autoimmune inflammatory eye disease; allergy; ss.
 XX

OS	Homo sapiens.
XX	WO200009552-A1.
XX	24-FEB-2000.
XX	13-AUG-1999; 99WO-US18298.
XX	14-AUG-1998; 98US-0096622.
PR	17-AUG-1998; 98US-0096815.
XX	04-SEP-1998; 98US-0099229.
PR	23-OCT-1998; 98US-0105368.
PR	08-JAN-1999; 99US-0115234.
PR	12-FEB-1999; 99US-0119931.
PR	18-FEB-1999; 99US-0120575.
PR	30-APR-1999; 99US-0132020.
PR	11-AUG-1999; 99US-0096622.
XX	(GEMY) GENETICS INST INC.
XX	Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI	Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI	Wong GG, Clark HF, Fechtel K;
DR	WPI: 2000-205979/18.
DR	P-PSDB; AA94973.
XX	New polynucleotides encoding secreted proteins, which may have e.g.
PT	nutritional, chemokine, immune stimulating or suppressing,
PT	hematopoiesis regulating, tissue growth, activin/inhibin
PT	antiinflammatory or tumor inhibition activity
XX	Claim 160; Page 606-607; 641pp; English.
XX	AA94618 to AA916697 encode the human secreted proteins given in
CC	AA946898 to AA94980, isolated from human adult brain, adult thyroid,
CC	adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC	adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC	foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC	and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC	predicted to have biological activities which would make them suitable
CC	for treating, preventing or ameliorating medical conditions in humans
CC	and animals. The polynucleotides can be used as markers for tissues in
CC	which the protein is preferentially expressed, as molecular weight
CC	markers on Southern gels, and as chromosome markers or tags to identify
CC	chromosomes or to map gene positions. The proteins can be used in the
CC	treatment of immune deficiencies and disorders, such as severe combined
CC	immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC	infections. These infections include human immunodeficiency virus (HIV),
CC	hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC	cardiailis. The proteins can be used to treat autoimmune disorders such
CC	as connective tissue disease, multiple sclerosis, systemic lupus
CC	erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC	Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC	diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC	autoimmune inflammatory eye disease. The proteins can also be used to
CC	treat allergic conditions, such as asthma. AA916698 to AA91674 represent
CC	probes for the human secreted proteins from the present invention.
XX	Sequence 1330 BP: 364 A; 285 C; 309 G; 372 T; 0 other;

Query Match 11.4%; Score 86.4; DB 21; Length 1330;
Best Local Similarity 51.1%; Pred. No. 1.7e-10;
Matches 232; Conservative 0; Mismatches 216; Indels 6; Gaps 1;

Qy	216	aaaaattattgctagaaaaaatgaaatactttagggactatccagatctctgtttgggaatta	275
Db	485	agaattcttgaaggaggagaccccaagaatcccttgggttggcagattctgacTgcacctga	544
Qy	276	tgaacctttctttgggagttatcttctcttttcactttgttaaacaccatatccaagtttc	335
Db	545	tgaaccttagcataggaataaacatgatgtgtatggcatctaatacttaTgggaagtaaac	604

FTXXPNXXPDXXPFXXPRXXPAXXPIXXDRXXPTXXPSXXCC

US5977072-A.
02-NOV-1999.
15-DEC-1998; 98US-0213389.
21-AUG-1997; 97US-0916902.
(INCYTE-) INCYTE PHARM INC.
Bandman O, Corley NC, Lal P;
WPI; 2000-012123/01.
P-PSDB; AAY50174.
High affinity immunoglobulin E receptor-like protein useful for
stimulating allergic and immune responses -
Example 5; Fig 1; 29pp; English.

570ctgttaactgtctctgttcttgggaattttgattacattgatgaacttcacgaattattgaat629
686ggactatgtccatcttaataatgggtctctggatggcatgggtgctctcttaagtgtgctggaat745
630tattcattctctgccttctctcaatttttgggggtg663
746tctgcatgtgtgtccctctctctgctcttggatg779

Search completed: August 30, 2002, 02:46:09
Job time: 4347 sec

FTXXPNXXPDXXPFXXPRXXPAXXPIXXDRXXPTXXPSXXCC

High affinity immunoglobulin E receptor-like protein useful for
stimulating allergic and immune responses -
Example 5; Fig 1; 29pp; English.

This sequence represents cDNA encoding human high affinity immunoglobulin
E (IgE) receptor-like protein (IGERB). Nucleic acids encoding IGERB were
initially identified in a brain cDNA library, this sequence being a
consensus. An allergic response is initiated by release of IgE
from B lymphocytes. The IgE molecules then bind to the high affinity
IgE receptor (FCRI) present on mast cells and basophils, which triggers
the release of histamine and proteases from cytoplasmic granules and
leads to the synthesis of effectors of the allergic and inflammatory
response, such as prostaglandins, leukotrienes and cytokines. As IGERB
binds IgE, it may be administered to stimulate allergic and immune
responses in patients in whom IGERB is under expressed or inactive and
to supplement the patients own production of the protein. IGERB may also
be used as an antigen for the production of antibodies and to identify
candidate agonists and antagonists of IGERB expression and activity.
The antibodies may also be used in diagnosis. Antibodies and antagonists
may be administered to downregulate IGERB activity and reduce the
potency of inflammatory and allergic responses. They may be used in this
way to treat inflammatory disorders such as multiple sclerosis,
osteoarthritis, asthma and some complications of cancer. Conversely, the
agonists may be used to enhance immune responses.

Sequence 1670 BP; 481 A; 350 C; 334 G; 501 T; 4 other;

Query Match11.4%; Score 86.4; DB 21; Length 1670;
Best Local Similarity51.1%; Pred. No. 1.8e-10;
Matches 232; Conservative0; Mismatches 216; Indels6; Gaps1;

QY216aaaaattattgtcagaaaaatgaaattcttagggactatccagatcctgtttggaatta275
DB326agaagtcttgaaggagaaaccaaagtccttgggtgtgagattctgactgccctga385
QY276tgacctttcttttgagttattctcttcttcaactttgttaaaacacatccaaagtttc335
DB386tgagccttagcatgggaataacaatgatgtatgcatcctaatactatggaagtaacc445
QY336ctttatattcttccagatattcattctgggctctgtttgttcatttaattctggag395
DB446ctatttcogtgatatcgggtgacacaaatttgggggtccagtaattgtttatttcaggat505
QY396ctctcctaattgcagtgaagaaacacacagaaactctgataattatgagccgaataa455
DB506ccttgcaattgcagcaggaattagaaactacaaaagcctgtccagagtagtctaggaa565
QY456tgaattttcttagtcgctggggaataagtcgggaatacttctctcacatttggttca515
DB566tgaatatcaccagctctgtactgctgcgcagggatcttaatacaacattagcttgg625
QY516tcttagat-----caaaactacatttgggttattctaccacaaaatagtcagtgtaagg569
DB626cgttttattccatccacccttactgttaactactatggcgaactcaataattgtcatg685

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 30, 2002, 00:42:07 ; Search time 1960.86 Seconds
(without alignments)
8110.813 Million cell updates/sec

Title: US-09-821-821-1
Perfect score: 760
Sequence: 1 ttccagtgtccaggcagcc.....aaaaaaaaaaaaaaaaaaaaa 760

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 1797656 seqs, 10463268293 residues 3595312
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg_inv.*
- 32: em.htg_other.*
- 33: em.htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description

1	760	100.0	760	6	AX268522	Sequence
2	689.2	90.7	694	9	AF237907	Homo sapi
3	687.4	90.4	697	9	AF321127	Homo sapi
4	686.2	90.3	691	6	AB013103	Homo sapi
5	666.4	87.7	676	6	AX179789	Sequence
6	598.2	78.7	603	6	AX179781	Sequence
7	377.2	49.6	382	6	AX335001	Sequence
8	347.4	45.7	450	6	AX179787	Sequence
9	340.2	44.8	417	6	AX179785	Sequence
10	250.4	32.9	138097	2	AC027787	Homo sapi
11	250.4	32.9	138097	2	AC027787	Homo sapi
12	250.4	32.9	147788	2	AC015840	Homo sapi
13	250.4	32.9	161238	9	AP003127	Homo sapi
14	250.4	32.9	166793	2	AC090401	Homo sapi
15	250.4	32.9	167934	9	AP001034	Homo sapi
16	242.8	31.9	246	6	AX179783	Sequence
17	88	11.6	1076	9	AF068288	Homo sapi
18	86.4	11.4	720	9	AB022821	Homo sapi
19	86.4	11.4	916	9	AB013102	Homo sapi
20	86.4	11.4	1550	9	BC020648	Homo sapi
21	86.4	11.4	1619	9	AF237912	Homo sapi
22	86.4	11.4	1669	6	AR035695	Sequence
23	86.4	11.4	1669	6	AR083580	Sequence
24	83.2	10.9	959	9	AF350500	Homo sapi
25	80.2	10.6	2805	9	AF367473	Homo sapi
26	64	8.4	1222	4	SSR236932	Sus scrofa
27	63.6	8.4	2997	9	AK057418	Homo sapi
28	62.6	8.2	522	9	AF354928	Homo sapi
29	61.8	8.1	708	6	AX101306	Sequence
30	61.8	8.1	1010	10	MUSFCERB	Mouse mast
31	61.6	8.1	1646	9	HUMIERB	Homo sapien
32	61.6	8.1	1661	6	AB06627	Sequence 3
33	61.6	8.1	1661	6	AR082014	Sequence
34	61.6	8.1	1661	6	BD003343	HT m4 , m
35	61.6	8.1	1661	6	I25809	Sequence 3
36	60	7.9	1628	9	BC008487	Homo sapi
37	58.2	7.7	1474	6	I07272	Sequence 19
38	58.2	7.7	1474	6	I07277	Sequence 24
39	58.2	7.7	1476	6	I07277	Human mRNA
40	58.2	7.7	1482	9	BC002807	Homo sapi
41	58.2	7.7	1597	6	AX333252	Sequence
42	58.2	7.7	1597	9	HSCD20A	Human mRNA
43	57.2	7.5	899	10	AF237910	Mus muscu
44	57	7.5	1163	10	AF237915	Mus muscu
45	55	7.2	1146	9	HUMBLIYM	Human B-lym

ALIGNMENTS

RESULT 1	AX268522	Sequence 1	from Patent WO0174903.	DNA	linear	PAT 29-OCT-2001
LOCUS	AX268522	Sequence 1	from Patent WO0174903.			
DEFINITION	AX268522	Sequence 1	from Patent WO0174903.			
ACCESSION	AX268522.1	GI:16541676				
VERSION	human.					
KEYWORDS	human.					
SOURCE	human.					
ORGANISM	Homo sapiens					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
REFERENCE	1 (sites)					
AUTHORS	Welcher, A.A. and Calzone, F.J.					
TITLE	Cd20/ige-receptor like molecules and uses thereof					
JOURNAL	Patent: WO 0174903-A 1 11-OCT-2001;					
FEATURES	Amgen, Inc. (US)					
source	Location/Qualifiers					
	1..760					
	/organism="Homo sapiens"					
	/db_xref="taxon:9606"					
	98..700					
CDS	/note="unnamed protein product"					
	/codon_start=1					

/protein_id="CAD10310.1"
/db_xref="GI:16541677"
/translation="MDSSTAHSPVFLVFPPEITASEVESPELSATTFSTQSPLOKILFA
RKMKILGTIQLFGIMFSGVIFLTLKPYPRPFIFLSGYPWGSVLFINSGLF
TAVKRKTTETLILSRINWFLSALGATAGIILLTGFILDONVIGYSHNSOCKAVT
VLFGILITLMTFSIIELFISLPSILGHSCEDCCEQC"
BASE COUNT 232 a 157 c 126 g 245 t
ORIGIN

Query Match 100.0%; Score 760; DB 6; Length 760;
Best Local Similarity 100.0%; Pred. No. 1.2e-161;
Matches 760; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ttccagtgtccaggcagctcagcagcaagaaagaaacatggtctagactgaagtaccaa 60
Db 1 TTCCAGTGTCTCCAGGAGCCTCAGCACAGAAAGAAACATGCTAGACTGAAGTACCAA 60
Qy 61 ctaaatcatctcttcaaatatcaccagaccatcatgattcagaccagccacacag 120
Db 61 CTAAATCATCTCCTTTCAAAATATCACCGACACCATCATGATTCAGACCCACACAG 120
Qy 121 tccggtgttcttggtatttctccagaatacactgcttcacagaatatgagtcacagaact 180
Db 121 TCCGGGTGTTTCTGGTATTTCCTCCAGAAATCACTGCTTCAGAAATATGAGTCCAGAACT 180
Qy 181 ttccagcagacactttcaactcaagcccttcgaaaaattatttctagaaaaatgaa 240
Db 181 TTCCAGCACGACCTTTCAACATCAAGCCCTTCGAAAATATTTCTAGAAAATGAA 240
Qy 241 aattcaggagactccagatcctgttggaaattatgacctttcttttggagttatctt 300
Db 241 AATCTTAGGACTATCCAGATCCTGTTGGAAATATGACCTTTCTTTTGGAGTTATCTT 300
Qy 301 cctttcactttgttaaaaccataccaggtttccctttatattttttcagatatcc 360
Db 301 CCTTTTCACTTTGTHAAACCAATATCCAGGTTTCCCTTTATATTTCTTCAGAGATCC 360
Qy 361 attctggggtctgtttgttcttaattctggagccttctcattgagtgaaagaaa 420
Db 361 ATTCTGGGGCTCTGTTTGTTCATTAATCTGGAGCCTTCCTAATTCAGTGAAAGAAA 420
Qy 421 aaccacagaactctgataattagccgaataatgaatttcttagtccctggagc 480
Db 421 AACCACAGAACTCTGATTAATATGACCGGAATATGAAATTTCTTAGTCCCTGGGAGC 480
Qy 481 aatagctggaatcattctcctcacatttgggtttcattcctagatcaaaactacattgtg 540
Db 481 AATAGCTGGAATCATTTCTCCTCACATTTGTTTTCATCTCCTAGATCAAAACTACATTTGTG 540
Qy 541 ttattctccaaaatagtcaggttaagcgtttactgtctctgtctcttgggaatttggat 600
Db 541 TTATTTCCACCAAAATAGTCAGTGAAGGCTGTACTGTCTCTGTCTTGGGAATTTTGTAT 600
Qy 601 tacattgagactttcagcattatgaattatcattctctgccttctcattttggg 660
Db 601 TACATTTGATGACTTTTTCAGCATTTATGAAATATTCATTTCTCTGCTTTCTCAATTTTGGG 660
Qy 661 gtccactcagaggttgattggaacatgattgtttgactagcactgtgagaataaag 720
Db 661 GTGCCACTCAGAGGATGTGATTGTGACAAATGTGTTGTGACTAGCAGTGTGAGAAATAAG 720
Qy 721 atgtgttaaatctcaaaaaaataaaaaaataaaaaa 760
Db 721 ATGTGTTAAATCTCAAAAAAATAAAAAAATAAAAAA 760

RESULT 2
AF237907 AF237907 694 bp mRNA linear PRI 17-APR-2001
LOCUS Homo sapiens MS4A5 protein mRNA, complete cds.
DEFINITION AF237907
ACCESSION AF237907
VERSION AF237907.1 GI:13649400

KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 694)
AUTHORS Liang, Y. and Tedder, T.F.
TITLE Identification of a cd20-, fcepsilonribeta-, and htm4-related gene family: sixteen new ms4a family members expressed in human and mouse
JOURNAL Genomics 72 (2), 119-127 (2001)
MEDLINE 21295030
REFERENCE 2 (bases 1 to 694)
AUTHORS Liang, Y. and Tedder, T.F.
TITLE Direct Submission
JOURNAL Submitted (24-FEB-2000) Immunology, Duke Medical Center, Research Dr., Durham, NC 27710, USA
FEATURES
Location/Qualifiers
source 1..694
/organism="Homo sapiens"
/db_xref="taxon:9606"
55..657
/codon_start=1
/product="MS4A5 protein"
/protein_id="AAK37416.1"
/db_xref="GI:13649401"
/translation="MDSSTAHSPVFLVFPPEITASEVESPELSATTFSTQSPLOKILFA
RKMKILGTIQLFGIMFSGVIFLTLKPYPRPFIFLSGYPWGSVLFINSGLF
TAVKRKTTETLILSRINWFLSALGATAGIILLTGFILDONVIGYSHNSOCKAVT
VLFGILITLMTFSIIELFISLPSILGHSCEDCCEQC"
BASE COUNT 198 a 143 c 116 g 237 t
ORIGIN
Query Match 90.7%; Score 689.2; DB 9; Length 694;
Best Local Similarity 99.6%; Pred. No. 1.1e-145;
Matches 691; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 44 ctgagctgaagtaccacataatcatctccttcaaaattatcacccagacacatcatgat 103
Db 1 CTAGACTGAAGTACCACATAAATCATCTCTTTCAAAATATCACCCGACACCATCATGGAT 60
Qy 104 tcaagcacgcacacagtcggtgtttctggtattctctccagaaatcactgctcagaa 163
Db 61 TCAGCACCACACACAGTCGGGTGTTCTGGGATTTCTCCAGAAATCAGTCCTTCAGAA 120
Qy 164 tatgagtcacagaactttcagccagacacttttcaactcaaaagcccttgcaaaaaa 223
Db 121 TATGAGTCCACAGAACTTTTCAGCCAGACCTTTTCACTCAAAAGCCCTTGCAAAAATTA 180
Qy 224 ttgtctagaaaaatgaaaaatcttagggactatccagatcctgtttggaattatgacctt 283
Db 181 TTGTCTAGAAAAATGAAATCTTAGGGACTATCCAGATCTCTTTGGAAATATGACCTTT 240
Qy 284 tctttggagttatcttctccttttccactttgttaaaaccatatccaaaggtttccctttata 343
Db 241 TCCTTTGGAGTTATCTCTCTTTTCACTTTGTTTAAAAACCATCAAGGTTTCCCTTTATA 300
Qy 344 ttctttcagagatatacattctcggggtctgttttggctcatttaattcttgagccttcta 403
Db 301 TTCTTTTCAGGATATCCATCTCGGGGCTCTGTTTGTTCATTAATTTCTGGAGCCCTCTCTA 360
Qy 404 attgagtgaaaaagaaaaaccagaaactctgataaattatgagccgaaataaataattt 463
Db 361 ATTGCAGTGAAAAAGAAAAACCACAGAAACTCTGTGATAATATTGAGCCGAATAATGAATTT 420
Qy 464 cttagtgccctgggagcaatagctggaatcattctccctcacatttggtttcatcctagat 523
Db 421 CTTTAGTCCCTGGGAGCAATAGCTGGAAATCATTTCTCCACATTTGGTTTTCATCCTTAGAT 480
Qy 524 caaaactacatttgggttattctctcaacaaaaatagtcagtgtaagggtgttactgtcctg 583
Db 481 CAAAACTACATTTGTGTTATTCTCACCACAAAATAGTCAGTGTGAAGGCTGTTACTGTCTCTG 540

Qy	162	aataatgagtcacagaactttcagccacgaccttttcaactcaaacgcccttgcaaaaat	222
Db	121	AATATGAGTCACACAGAACTTTTCAAGCCACGACCCTTTTCACTCAAAGCCCCTTGCAAAAT	180
Qy	222	tattgtcacaaaaataaaactcttagggactaccagatccctgtttggaattatgatcct	281
Db	181	TATTGTGTAGAAAATAAGAAATCTTAGGGACTATCAGATGCCCTGTGTGGNAATTATGCACCT	240
Qy	282	ttctctttggagtatatctctcttttcaactttgtttaaacaccatcatcaagggtttcccctta	341
Db	241	TTTTCTTTTGGAGTTATCTTCTCTTTTCACTTTGTGTAAAAACCATATCCAAGGTTTCCCTTTTA	300
Qy	342	tatttcttcaggatatccaattctgggggcctgttttgttttcattaataattctggagccttcc	401
Db	301	TATTTCCTTTTCAGGATATCCCATCTCGGGGCTCTGTGTGTTCATTAATCTCGAGCCCTTCC	360
Qy	402	taattgcagtgaagaagaaaaaacacacagaaaactctgataaatattgagccgaataatgaatt	461
Db	361	TAATTCAGTGAANAAGAAAACACACAGAAACTCTGATAAATTATTGAGCCGAATAATGAATT	420
Qy	462	ttcttagtgccttggggagcaatagctggaatcatctctctccatcttggtttcatctoctag	521
Db	421	TTCTTAGTGCCTGGGAGCAATAGCTGGAATCATCTCTCCCTCACATTTGGTTTTCATCCTAG	480
Qy	522	atcaaaaactaattgttgattattctcacaataatgactgtagtgtaaggctgttactgtcc	581
Db	481	ATCAAAACTACATTTGTGGTTATTCTCACAAAATAGTCAGTGTAGGCTGTGTACTGTCTCC	540
Qy	582	tgctctgggaattttgattacattgatgactttccagcattattgaattattcattcttc	641
Db	541	TGTTGTGGGAATTTTGATTACATTGATGACTTTCACCATTAATTGAATTATTCATTCTTC	600
Qy	642	tgctctttotcaattttggvggtgcacotcagaggattgtgattgtgaacaaatgttgttgac	701
Db	601	TGCCTTCTCAATTTTGGGVTGCCACTCAGAGGATTTGTGATTGTGAACAATATTGTTGTGAC	660
Qy	702	tagcacgtgagaataaagatgtgttaaacttccaaa	738
Db	661	TAGCACGTGAGAATAAGATGTGTTAAAAATAAAAA	697
RESULT	4		
AB013103			
LOCUS	AB013103	691 bp	linear
DEFINITION	Homo sapiens mRNA for MS4A5, complete cds.		PRI 20-MAR-2000
ACCESSION	AB013103		
VERSION	AB013103.1	GI:11559213	
KEYWORDS	MS4A5; CD20-like 2.		
SOURCE	Homo sapiens testis cdna to mRNA.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (sites)		
AUTHORS	Ishibashi,K., Suzuki,M., Sasaki,S. and Imai,M.		
TITLE	Identification of a new multigene four-transmembrane family (M) related to CD20, HTm4 and beta subunit of the high-affinity Ig receptor		
JOURNAL	Gene 264 (1), 87-93 (2001)		
MEDLINE	21142397		
REFERENCE	2 (sites)		
AUTHORS	Ishibashi,K., Sasaki,S. and Marumo,F.		
TITLE	Cloning of three CD20 homolog from human, putative calcium channel		
JOURNAL	Unpublished		
REFERENCE	3 (bases 1 to 691)		
AUTHORS	Ishibashi,K.		
TITLE	Direct Submission		
JOURNAL	Submitted (20-APR-1998) Kenichi Ishibashi, Tokyo Medical and Dental University, 2nd Internal Medicine; Yushima 1-5-45, Bunkyo, Tokyo 113-8519, Japan (E-mail:kishibashi.med2@med.tmd.ac.jp, Tel:81-3-5803-5223, Fax:81-3-5803-0132)		
FEATURES	Location/Qualifiers		
source	1..691		

/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="testis"
52..654
/gene="MS4A5"
52..654
/gene="MS4A5"
/codon_start=1
/product="MS4A5"
/protein_id="BAB18739.1"
/db_xref="GI:11559214"
/translation="MSSSTAHSVPFLVPPPEITASVESTELSATTFTSQSPLOKFLA
RKMKILGIIQLIGIMTFSEGVIFLFLKPKYPRFPFIPLSGYPFGWSVLFNSGAFI
IAYRKTEFLIILSMNPLSALGAIILITFGILFDQNYICGYSQNSQCKAVT
VLFILGILITLMTLSIIEFLSFLPSILGHSIEDCDCEQC"
BASE COUNT 195 a 142 c 117 g 237 t
ORIGIN

Query Match 90.3%; Score 686.2; DB 9; Length 691;
Best Local Similarity 99.6%; Pred. No. 5.4e-145;
Matches 688; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 47 gactgaagtaaccactaaatcatctcttccaaattatcccgacacacacatcgattcca 106
Db 1 GACTGAAGTACCAACTAAGTCATCTCTTCAAATATATCACCGACACCATCGGATCCA 60
Qy 107 agcaccgcacacagtcgggtgttctggtatttctccagagaatcactgcttcagaatat 166
Db 61 AGCACCGCACACATCGGGTGTCTGATTTCTCCAGAAATCAGTGTTCAGAAATAT 120
Qy 167 gactccagagaacttcagccagacacacttttcaactcaaaagcccttgcaaaaattatt 226
Db 121 GAGTCCACAGAACTTCAGCCACGACCTTTCACTCAAAAGCCCTTGCAAAAATATTT 180
Qy 227 gctgaagaaatgaaatcttaggactatccagatcctgtttgagaattatgacatttct 286
Db 181 GCTAGAAATGAAATCTTAGGAGACTATCCAGATCTGTTTGGAAATATGACCTTTCT 240
Qy 287 ttggaggtattcttcttccactttgttaaaccatatccaaagggttccctttatatt 346
Db 241 TTTGGAGTTATCTCTCTTTTCACTTCTTAAACCATATCCAAGGTTTCCCTTTATATT 300
Qy 347 ctctcagatatcattctagggtctgtttgttcaattaatctcaggagcttcctaat 406
Db 301 CTTTCAGGATATCATTTGGGGCTCTGTTTGTTCATTAAATCTGGAGCCTTCTCTAAAT 360
Qy 407 gcagtgaagaaagaaacacagaaactctgataattatgagccgaataatgaaatttctt 466
Db 361 CAGTGAAGAAAGAAACACAGAAACTCTGATAATATTGAGCCGGAATATGAAATTTCTT 420
Qy 467 agtgcctgggagcaatagctgggaatcattctcctcacatttgggttccatccatgata 526
Db 421 AGTCCCTGGGAGCAATAGCTGGAATCATCTCCTCACATTTGGTTTCATCCTTAGATCAA 480
Qy 527 aactacatttgggttattctcaccaaaatagtcagtgtaagcgttactgctcctgttc 586
Db 481 AACTACATTTGGGTATTCTCACCAAAATAGTCAGTGTAAAGCCTGTACTGTCCCTGTTTC 540
Qy 587 ttgggaatttggattacattgatgactttcagcattattgaattattcatttctctgct 646
Db 541 TTGGGAATTTGTATTACATTGATGACTTTTCAGCATTTATTGAAATATTCTCTCTGCTT 600
Qy 647 ttctcaatttgggggtgccactcagaggattgtgattggaacaattgttggtagactagca 706
Db 601 TTCTCAATTTTGGGGTGCCTACTCAGAGGATTTGTGATTGTGAACAAATGTTGTTGACTAGCA 660
Qy 707 ctgtgagaataaagatgtgttaaaatctcaa 737
Db 661 CTGTGAGAAATAAGATGTGTTAAATATAAA 691

RESULT 5

AX179789
LOCUS AX179789 676 bp DNA linear PAT 06-AUG-2001
DEFINITION Sequence 9 from Patent WO0146417.
ACCESSION AX179789
VERSION AX179789.1 GI:15132148
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Walke,D.W. and Turner,C.A.
JOURNAL Human membrane proteins and polynucleotides encoding the same
having homology to cd20 proteins and ige receptors
Patent: WO 0146417-A 9 28-JUN-2001;
Lexicon Genetics Incorporated (US)
FEATURES
Location/Qualifiers
1..676
/Organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 193 a 140 c 112 g 231 t
ORIGIN

Query Match 87.7%; Score 666.4; DB 6; Length 676;
Best Local Similarity 99.1%; Pred. No. 1.6e-140;
Matches 670; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 66 tcattctcttcaaaattatcccgacacacatcatgattccaaagccacacacagtcgg 125
Db 1 TCATCTCTCTTTCAAATATATCACCGACACCATCATGGATTCAAGCACCCGACACATCGCG 60
Qy 126 tgttctggtatttctccagaaatcactgcttcagaatatgagtcacacacactttcag 185
Db 61 TGTCTCTGATTTCTCTCCAGAAATCAGTGTTCAGAAATATGAGTCCACAGAACTTTCCAG 120
Qy 186 ccacgaccttttcaactcaaaagcccttgcaaaaattatttgcagaaatgaaaatct 245
Db 121 CCACGACCTTTTCAAATCAAAAGCCCTTGCAAAAATATTATTTGCTAGAAAATGAAAATCT 180
Qy 246 taggactatccagatcctgttggaattatgacatttcttcttgagattatcttctt 305
Db 181 TAGGACTATCCAGATCTGTTTGGAAATATGACCTTCTTTTGGAGTTATCTTCTCTT 240
Qy 306 tcactttgttaaaacatatccaaaggttccctttatatttcttccagatatccattct 365
Db 241 TCACCTTGTTAAACCATATCCAAGTTTCCCTTTATATTCTTTTCAGGATATCCATTCT 300
Qy 366 ggggtctgtttgttcttaattctcaggagcttctcctaatcagtgagtgaaagaaacca 425
Db 301 GGGGCTCTGTTTGTTCATTAAATCTGGAGCCTTCTTAATTCAGTGAAGAAAGAAACCA 360
Qy 426 cagaactcgaataattgagccgaataatgaaatttcttagtgccctgggagcaatag 485
Db 361 CAGAAACCTCTGATAATATGAGCCGAATAATGAATCTTCTTAGTGCCCTGAGAGCAATAG 420
Qy 486 ctggaatcattctcctcacatttgggttccatccatagacaaactacatttgggttatt 545
Db 421 CTGGAATCATCTCCTCACATTTGGTTTTCATCCINAGATCAAACTACATTTTGGTTATT 480
Qy 546 ctcaacaaaatagtcagtgtaagcgtgtaactgtctgttcttgggaaatttggattacat 605
Db 481 CTCACAAAATAGTCAGTGTAAAGCCTGTACTGTCTCTTCTTGGAAATTTTGATTACAT 540
Qy 606 tgatgacttcagcattattgaattattcatttctctgccttctcatttcttgggggccc 665
Db 541 TGATGACTTTTACGCAATATTGAATTTATTTCTCTGCCCTTCTCAATTTTGGGGTCCC 600
Qy 666 actcagaggattgtgattgtaacaaatgtgtgactagcactgtgagaataaaagattg 725
Db 601 ACTCAGAGGATTTGATTGTGAACAAATTTGTTGACTAGCAGTGTGAGAAATAAGATGTG 660
Qy 726 ttaaaatctcaaaaaa 741

```
||||||| |||||
Db 661 TTAATAAATAAAAAA 676

RESULT 6
AX179781
LOCUS AX179781 603 bp DNA linear PAT 06-AUG-2001
DEFINITION Sequence 1 from Patent WO0146417.
ACCESSION AX179781
VERSION AX179781.1 GI:15132144
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 603)
AUTHORS Walke,D.W. and Turner,C.A.
TITLE Human membrane proteins and polynucleotides encoding the same
having homology to cd20 proteins and ige receptors
JOURNAL Patent: WO 0146417-A 1 28-JUN-2001;
Lexicon Genetics Incorporated (US)
FEATURES
Location/Qualifiers
source 1..603
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 162 a 125 c 104 g 212 t
ORIGIN

Query Match 78.7%; Score 598.2; DB 6; Length 603;
Best Local Similarity 99.5%; Pred. No. 4e-125;
Matches 600; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 98 atgattcaaacgacacacagcccggtgtttctgtgtattctccagaaatcaactgct 157
Db 1 ATGATTCAACACGACGACAGTCCGGTGTTCGTGTATTCTCTCAGAAATCAGTCT 60

Qy 158 tcagaatatgagtcacagaacttccagccagccactttccaactcaaaaccccttgc 217
Db 61 TCAGAAATATGAGTCACAGAACTTTCAGCCAGCACCTTTCAACTCAAAAGCCCTTGC 120

Qy 218 aaattatttgcagaaatgaaactcttaggaactccagactccctgtttgaaattatg 277
Db 121 AAATATTGTGAGAAATGAAATCTTAGGGACTATCCAGATCCCTGTTTGGAAATATG 180

Qy 278 acctttcttggagtattctctcttccactttgttaaacaatccaaaggtttccc 337
Db 181 ACCTTTCTTTGGAGTATCTCTCTTTTCACCTTGTATAAACCATATCAAGGTTTCCC 240

Qy 338 ttatattcttccaggatatacattctggggtctgttttggttcatttaattctggagcc 397
Db 241 TTTATATTCTTTGAGGATATCCATTTCTGGGGCTCTGTTTGTTCATTAATTTCTGGAGCC 300

Qy 398 ttcctaattgcagtgaaagaaacacacagaaactctgataatattgagccgaataatg 457
Db 301 TTCCTAATGTCAGTGAAAGAAACACAGAAACTCTGATAATATTATGAGCCGAAATATG 360

Qy 458 aattttcttagtcctgggagcaatagctggaatcatctctccacatttggtttccatc 517
Db 361 AATCTTCTTAGTGCCCTGAGAGCAATAGCTGGAATCATCTTCCTCCACATTTGGTTTCATC 420

Qy 518 ctagataaaactacatttgggtattctcaccaaaatagtcagtgtaaggctgttact 577
Db 421 CTAGATCAAAACTACATTTGGTGTATTCTCACCAAAATAGTCAGTGAAGGCTGTACT 480

Qy 578 gtccgttcttggaaatttgaatcacattgatgactttcagcattattgaatttcatt 637
Db 481 GTCCCTGTTCTTGGGAATTTTGATACATTGATGACTTTCAGCATATTGAATTAATTCATT 540

Qy 638 tctctgcctttctcaattttgggtggccactcagaggtattgtattgtgaacaattgtgt 697
Db 541 TCTCTGCCTTCTCAATTTTGGGGTGCCACTCAGAGGATTGTGATTGTGAACAATGTGT 600
```

```
Qy 698 tga 700
Db 601 TGA 603

RESULT 7
AX335001/c
LOCUS AX335001 382 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 5510 from Patent WO0194629.
ACCESSION AX335001
VERSION AX335001.1 GI:18125720
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
Horrigan,S., Soppet,D.R. and Weaver,Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 5510 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
Location/Qualifiers
source 1..382
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 135 a 73 c 67 g 107 t
ORIGIN

Query Match 49.6%; Score 377.2; DB 6; Length 382;
Best Local Similarity 99.2%; Pred. No. 3.2e-75;
Matches 379; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 356 tatcaattctgggctctgtttgttcatttaattctggagccttctaattgcagtgaaa 415
Db 382 TATCCATTCTGGGGCTCTGTTTGTTCATTAATCTGGAGCCCTCTTAATTCAGTGAAA 323

Qy 416 agaaaaaccacagaaactctgataattaggccgaataataatttcttagtgccttg 475
Db 322 AGAAAAACCACAGAAACTCTGATAATATTAGCCGAATAATGAATTTCTTAGTGCCCTG 263

Qy 476 ggaagaatagctggaatactctcctcacatttggtttctcctagatcaaaactacatt 535
Db 262 GGAGCAATAGCTGGAATCATTTCTCCATTTGGTTCATCTAGATCAAAACTACATT 203

Qy 536 tfgtggtattctcaccaaaatagtcagtgtaaggctgttactgtctctgtcttgggaatt 595
Db 202 TGTGGTATTCTCACCAAAATAGTCAGTGTAGGCTGTACTGTCTGTCTTGGGAATT 143

Qy 596 ttgattacattgatgactttcagcattattgaattattcattctctctgcttctcaatt 655
Db 142 TTGATTACATTGATGACTTTCAGCAATTATTGAATTTATTCTTCTGCTTTCTCAATT 83

Qy 656 ttgggtgcactcagagattgtgattgaaacaatgttggtagcactagcactgtgagaa 715
Db 82 TTGGGTGCCACTCAGAGGATTTGATTGTGAACAATTTGTGTGACTAGCAGCTGTGAGAA 23

Qy 716 taaagatgtgttaaaatctcaa 737
Db 22 TAAAGATGTGTAAATATATAA 1

RESULT 8
AX179787
LOCUS AX179787 450 bp DNA linear PAT 06-AUG-2001
DEFINITION Sequence 7 from Patent WO0146417.
ACCESSION AX179787
VERSION AX179787.1 GI:15132147
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 450)

REFERENCE
AUTHORS

Walke,D.W. and Turner,C.A.
Human membrane proteins and polynucleotides encoding the same
having homology to cd20 proteins and ige receptors

JOURNAL

Patent: WO 0146417-A 5 28-JUN-2001;
Lexicon Genetics Incorporated (US)
Location/Qualifiers

FEATURES

source
1. .450
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 120 a 94 c 77 g 159 t

ORIGIN

Query Match 45.7% Score 347.4; DB 6; Length 450;
Best Local Similarity 97.0% Pred. No. 1.7e-68;
Matches 354; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 98 atgattcaagcaccgacacacagtcggtttctgtgtattctccagaaatcaactgct 157
|||||

Db 1 ATGGATTCAAGCACCACACAGTCGGTGTCTCTGTGTAATTCCTCCAGAAATCACTGCT 60
|||||

Qy 158 tcagaatatgagtcacacagaaatttcagccagacatttcaactcaaaagcccttgcaa 217
|||||

Db 61 TCAGAAATGAGTCCACAGAACTTTCAGCCAGCAGCCTTTTCAACTCAAAAGCCCTTGCAA 120
|||||

Qy 218 aaattattgctagaaaaatgaaattcttagggactatccagatcctgtttggaattatg 277
|||||

Db 121 AAATTATTGCTAGAAAATGAAATCTTAGGGACTATCCAGATCCTGTGTAATATG 180
|||||

Qy 278 acctttctttggagtattctcttcttcaactttgtttaaaccatatccaaggtttccc 337
|||||

Db 181 ACCTTTCTTTGGAGTTATCTTCTTTACCTTGTGTTAAACCATATCCAAAGGTTTCCC 240
|||||

Qy 338 ttatattcttcagatatccattctgggtctgtttgttgcatttaattcttgagcc 397
|||||

Db 241 TTTATATTTCTTCAGGATATCCATCTCGGGCTCTGTTTGTTCATTAATTCGAGCC 300
|||||

Qy 398 ttctaatgtcagtgaaagaaacacacagaaactctgataattgagccgaataatg 457
|||||

Db 301 TTCCTAATTCAGTGAAAGAAACCCACAGAACTCGGAATTTGATTACATTGATG 360
|||||

Qy 458 aattt 462
|||||

Db 361 ACTTT 365

RESULT 9

AX179785

LOCUS

Sequence 5 from Patent WO0146417. linear PAT 06-AUG-2001

ACCESSION

AX179785

VERSION

AX179785.1

KEYWORDS

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

Walke,D.W. and Turner,C.A.

TITLE

Human membrane proteins and polynucleotides encoding the same

JOURNAL

Patent: WO 0146417-A 5 28-JUN-2001;
Lexicon Genetics Incorporated (US)
Location/Qualifiers

FEATURES

source
1. .417
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 115 a 91 c 69 g 142 t

ORIGIN

Query Match 44.8% Score 340.2; DB 6; Length 417;
Best Local Similarity 99.1% Pred. No. 7.3e-67;
Matches 342; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 98 atgattcaagcaccgacacacagtcggtttctgtgtattctccagaaatcaactgct 157
Db 1 ATGGATTCAAGCACCACACAGTCGGTGTCTCTGTGTAATTTCTCCAGAAATCACTGCT 60
|||||

Qy 158 tcagaatatgagtcacacagaaatttcagccagacatttcaactcaaaagcccttgcaa 217
Db 61 TCAGAAATGAGTCCACAGAACTTTCAGCCAGCAGCCTTTTCAACTCAAAAGCCCTTGCAA 120
|||||

Qy 218 aaattattgctagaaaaatgaaattcttagggactatccagatcctgtttggaattatg 277
Db 121 AAATTATTGCTAGAAAATGAAATCTTAGGGACTATCCAGATCCTGTGTAATATG 180
|||||

Qy 278 acctttctttggagtattctcttcttcaactttgtttaaaccatatccaaggtttccc 337
Db 181 ACCTTTCTTTGGAGTTATCTTCTTTACCTTGTGTTAAACCATATCCAAAGGTTTCCC 240
|||||

Qy 338 ttatattcttcagatatccattctgggtctgtttgttgcatttaattcttgagcc 397
Db 241 TTTATATTTCTTCAGGATATCCATCTCGGGCTCTGTTTGTTCATTAATTCGAGCC 300
|||||

Qy 398 ttctaatgtcagtgaaagaaacacacagaaactctgataattgagccgaataatg 442
Db 301 TTCCTAATTCAGTGAAAGAAACCCACAGAACTCTGATCAAA 345
|||||

RESULT 10
AC027787
LOCUS
DEFINITION Homo sapiens chromosome 15 clone RP11-196E16 map 15, LOW-PASS
SEQUENCE SAMPLING.

ACCESSION AC027787
VERSION AC027787.2 GI:9845160
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 138097)
2 (bases 1 to 138097)
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

1 (bases 1 to 138097)
2 (bases 1 to 138097)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campoliano,A., Castle,A., Choquel,Y., Colangelo,M., Collins,S.,
Collins,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S.,
Dodgson,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieuw,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Murphy,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
O'Neill,D., Ollivar,T.M., Oliver,J., Peterson,K., Piere,N.,
Pisan,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye.W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission
Submitted (01-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 18, 2000 this sequence version replaced g1:7382629.
All repeats were identified using RepeatMasker.

Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9008
Center clone name: 196_E_16

* NOTE: This record contains 163 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 663: contig of 663 bp in length
* 664 763: gap of 100 bp
* 764 1464: contig of 701 bp in length
* 1465 1564: gap of 100 bp
* 1565 2268: contig of 704 bp in length
* 2269 2368: gap of 100 bp
* 2369 3042: contig of 674 bp in length
* 3043 3142: gap of 100 bp
* 3143 3820: contig of 678 bp in length
* 3821 3920: gap of 100 bp
* 3921 4626: contig of 706 bp in length
* 4627 4726: gap of 100 bp
* 4727 5423: contig of 697 bp in length
* 5424 5523: gap of 100 bp
* 5524 6234: contig of 711 bp in length
* 6235 6334: gap of 100 bp
* 6335 7040: contig of 706 bp in length
* 7041 7140: gap of 100 bp
* 7141 7837: contig of 697 bp in length
* 7838 7937: gap of 100 bp
* 7938 8632: contig of 695 bp in length
* 8633 8732: gap of 100 bp
* 8733 9430: contig of 698 bp in length
* 9431 9530: gap of 100 bp
* 9531 10214: contig of 684 bp in length
* 10215 10314: gap of 100 bp
* 10315 10980: contig of 666 bp in length
* 10981 11080: gap of 100 bp
* 11081 11783: contig of 703 bp in length
* 11784 11883: gap of 100 bp
* 11884 12582: contig of 699 bp in length
* 12583 12682: gap of 100 bp
* 12683 13389: contig of 707 bp in length
* 13390 13489: gap of 100 bp
* 13490 14186: contig of 697 bp in length
* 14187 14286: gap of 100 bp
* 14287 14996: contig of 710 bp in length
* 14997 15096: gap of 100 bp
* 15097 15796: contig of 700 bp in length
* 15797 15896: gap of 100 bp
* 15897 16583: contig of 687 bp in length
* 16584 16683: gap of 100 bp
* 16684 17376: contig of 693 bp in length
* 17377 17476: gap of 100 bp
* 17477 18178: contig of 702 bp in length
* 18179 18278: gap of 100 bp
* 18279 18962: contig of 684 bp in length
* 18963 19062: gap of 100 bp
* 19063 19673: contig of 611 bp in length
* 19674 19773: gap of 100 bp
* 19774 20436: contig of 663 bp in length

* 20437 20536: gap of 100 bp
* 20537 21233: contig of 697 bp in length
* 21234 21333: gap of 100 bp
* 21334 22042: contig of 709 bp in length
* 22043 22142: gap of 100 bp
* 22143 22808: contig of 666 bp in length
* 22809 22908: gap of 100 bp
* 22909 23595: contig of 687 bp in length
* 23596 23695: gap of 100 bp
* 23696 24391: contig of 696 bp in length
* 24392 24491: gap of 100 bp
* 24492 25190: contig of 699 bp in length
* 25191 25290: gap of 100 bp
* 25291 25981: contig of 691 bp in length
* 25982 26081: gap of 100 bp
* 26082 26765: contig of 684 bp in length
* 26766 26865: gap of 100 bp
* 26866 27563: contig of 698 bp in length
* 27564 27663: gap of 100 bp
* 27664 28377: contig of 714 bp in length
* 28378 28477: gap of 100 bp
* 28478 29184: contig of 707 bp in length
* 29185 29284: gap of 100 bp
* 29285 29983: contig of 699 bp in length
* 29984 30083: gap of 100 bp
* 30084 30775: contig of 692 bp in length
* 30776 30875: gap of 100 bp
* 30876 31565: contig of 690 bp in length
* 31566 31665: gap of 100 bp
* 31666 32360: contig of 695 bp in length
* 32361 32460: gap of 100 bp
* 32461 33157: contig of 697 bp in length
* 33158 33257: gap of 100 bp
* 33258 33951: contig of 694 bp in length
* 33952 34051: gap of 100 bp
* 34052 34735: contig of 684 bp in length
* 34736 34835: gap of 100 bp
* 34836 35483: contig of 648 bp in length
* 35484 35583: gap of 100 bp
* 35584 36308: contig of 725 bp in length
* 36309 36408: gap of 100 bp
* 36409 37108: contig of 700 bp in length
* 37109 37208: gap of 100 bp
* 37209 37892: contig of 684 bp in length
* 37893 37992: gap of 100 bp
* 37993 38672: contig of 680 bp in length
* 38673 38772: gap of 100 bp
* 38773 39475: contig of 703 bp in length
* 39476 39575: gap of 100 bp
* 39576 40290: contig of 715 bp in length
* 40291 40390: gap of 100 bp
* 40391 41087: contig of 697 bp in length
* 41088 41187: gap of 100 bp
* 41188 41891: contig of 704 bp in length
* 41892 41991: gap of 100 bp
* 41992 42679: contig of 688 bp in length
* 42680 42779: gap of 100 bp
* 42780 43500: contig of 721 bp in length
* 43501 43600: gap of 100 bp
* 43601 44295: contig of 695 bp in length
* 44296 44395: gap of 100 bp
* 44396 45088: contig of 693 bp in length
* 45089 45188: gap of 100 bp
* 45189 45880: contig of 692 bp in length
* 45881 45980: gap of 100 bp
* 45981 46683: contig of 703 bp in length
* 46684 46783: gap of 100 bp
* 46784 47481: contig of 698 bp in length
* 47482 47581: gap of 100 bp
* 47582 48290: contig of 709 bp in length
* 48291 48390: gap of 100 bp
* 48391 49077: contig of 687 bp in length
* 49078 49177: gap of 100 bp

```

* 49178 49879: contig of 702 bp in length
* 49880 49979: gap of 100 bp
* 49980 50684: contig of 705 bp in length
* 50685 50784: gap of 100 bp
* 50785 51507: contig of 723 bp in length
* 51508 51607: gap of 100 bp
* 51608 52320: contig of 713 bp in length
* 52321 52420: gap of 100 bp
* 52421 53133: contig of 713 bp in length
* 53134 53233: gap of 100 bp
* 53234 53933: contig of 700 bp in length
* 53934 54033: gap of 100 bp
* 54034 54703: contig of 670 bp in length

Query Match 32.9%; Score 250.4; DB 2; Length 138097;
Best Local Similarity 97.7%; Pred. NO. 1.2e-46;
Matches 254; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ttccagtgtccaggcagcctcagcacagaagaaagacatggtctagactgaagtaccaa 60
|||||
DB 109499 TTCCAGTGTCTCCAGGCGAGCTCAGCACAGAAAAGAAATGCTGTAGACTGAATACCAA 109558
|||||

QY 61 ctaaatcatctcttcaaatattaccgacaccatcatggtattcaagcaccgcacacag 120
|||||
DB 109559 CTAATCATCTCTTCAAAATATCACCACACACCATCATGATGATTCAAGCAGCCGACACAG 109618
|||||

QY 121 tccggtgttctggtatttctccagaatacactcttcagaataatgattccacagaact 180
|||||
DB 109619 TCCGGTGTCTCTGATTTCTCCAGAAATCACTGCTTCAGNATATGATTCACAGAACT 109678
|||||

QY 181 ttccgacagacaccttttcaactcaagcccttgcacaaataattttgtagaaaaatgaa 240
|||||
DB 109679 TTCCAGCACAGACCTTTTCAACTCAAGAGCCCTTGCAAAAATATTTGCTAGAAAAATGAA 109738
|||||

QY 241 aatcttaggactatccaga 260
|||||
DB 109739 AATCTTAGGGGTAAAGTAAGA 109758
|||||

RESULT 11
AC027787/c 138097 bp DNA linear HTG 18-AUG-2000
LOCUS Homo sapiens chromosome 15 clone RP11-196E16 map 15, LOW-PASS
DEFINITION SEQUENCE SAMPLING.
AC027787
VERSION AC027787.2 GI:9845160
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 138097)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-196E16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 138097)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,Y., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Holland,J.C., Illiev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klewin,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Menus,L., Mihova,T., Miranda,C., Mlenda,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,

```

Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
 Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (01-APR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 18, 2000 this sequence version replaced gi:7382629.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research

Center code: W1BR
 Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L9008

Center clone name: 196_E_16

 * NOTE: This record contains 163 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

* 1 663: contig of 663 bp in length
 * 664 763: gap of 100 bp
 * 764 1464: contig of 701 bp in length
 * 1465 1564: gap of 100 bp
 * 1565 2268: contig of 704 bp in length
 * 2269 2368: gap of 100 bp
 * 2369 3042: contig of 674 bp in length
 * 3043 3142: gap of 100 bp
 * 3143 3820: contig of 678 bp in length
 * 3821 3920: gap of 100 bp
 * 3921 4626: contig of 706 bp in length
 * 4627 4726: gap of 100 bp
 * 4727 5423: contig of 697 bp in length
 * 5424 5523: gap of 100 bp
 * 5524 6234: contig of 711 bp in length
 * 6235 6334: gap of 100 bp
 * 6335 7040: contig of 706 bp in length
 * 7041 7140: gap of 100 bp
 * 7141 7837: contig of 697 bp in length
 * 7838 7937: gap of 100 bp
 * 7938 8632: contig of 695 bp in length
 * 8633 8732: gap of 100 bp
 * 8733 9430: contig of 698 bp in length
 * 9431 9530: gap of 100 bp
 * 9531 10214: contig of 684 bp in length
 * 10215 10314: gap of 100 bp
 * 10315 10980: contig of 666 bp in length
 * 10981 11080: gap of 100 bp
 * 11081 11783: contig of 703 bp in length
 * 11784 11883: gap of 100 bp
 * 11884 12582: contig of 699 bp in length
 * 12583 12682: gap of 100 bp
 * 12683 13389: contig of 707 bp in length
 * 13390 13489: gap of 100 bp
 * 13490 14186: contig of 697 bp in length
 * 14187 14286: gap of 100 bp
 * 14287 14996: contig of 710 bp in length
 * 14997 15096: gap of 100 bp
 * 15097 15796: contig of 700 bp in length
 * 15797 15896: gap of 100 bp
 * 15897 16583: contig of 687 bp in length
 * 16584 16683: gap of 100 bp


```

Db 133619 TCCGGTGTTCCTCAGAAATCACTGCTTCAGAAATAGTCCACAGAACT 133678
QY 181 ttccagcagacctttcaaacctgcaaaatatttgcctagaaaaatgaa 240
|||||
Db 133679 TTCAGCCAGACCTTTCACTCAAGCCCTGCAAAATATTATTGCTAGAAAAGAA 133738
QY 241 aatctaggagctatccaga 260
|||||
Db 133739 AATCTTAGGGTAAGTAAGA 133758

RESULT 14
AC090401/1
LOCUS AC090401 166793 bp DNA linear HTG 14-OCT-2001
DEFINITION Homo sapiens chromosome 11 clone RP11-729B4 map 11, *** SEQUENCING
IN PROGRESS ***, 1 ordered pieces.
AC090401
AC090401.3 GI:15706110
VERSION HTG: HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 166793)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 11, clone RP11-729B4
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 166793)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,L., Boguslavsky,L., Boukhaltier,B., Brown,A.,
Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
Collins,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Fero,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., Lacombe,K., Lamazares,K., Landers,T.,
Lehocky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheeters,R., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
Roy,A., Santos,R., Schauer,S., Schuppback,R., Seaman,S., Severy,P.,
Sougnuez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 20, 2001 this sequence version replaced gi:14971278.
All repeats were identified using RepeatMasker:
Smith, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L12709
Center clone name: 729_B_4
-----
* NOTE: this is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and

```

```

* the accession number will be preserved.
*
FEATURES
source
1. 166793 bp in length.
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="RP11-729B4"
/clone_lib="pCI-11 Human Male BAC"
BASE COUNT 46459 a 31834 c 32914 g 55586 t
ORIGIN

Query Match 32.9%; Score 250.4; DB 2; Length 166793;
Best Local Similarity 97.7%; Pred. No. 1.1e-46;
Matches 254; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ttccagtgctcagcagccctcagcacacagaaaaaacatggtctagactgaagtaccaa 60
|||||
Db 144170 TTCCAGTGTCTCCAGCGACGCTCAGCACAGAAAAGAACATGGTCTAGACTGAAGTACCAA 144111
QY 61 ctcaatcatctcccttcaaaattatcacccagacaccatcatgattcgaagcacgcacacag 120
|||||
Db 144110 CTAATCATCTCCTTTCAAAATTATCACCGACACCATCATGATTCAGCACGCACACAG 144051
QY 121 tccggtgttctggtatttctccagaatacactgcttcagaatatgagtcacagaact 180
|||||
Db 144050 TCCGGTGTTCGTGATTTCCTCCAGAAATCACTGCTTCAGAAATATGAGTCCACAGACT 143991
QY 181 ttccagcagacctttcaactcaaacctgctgcaaaaattatttgcctagaaaaatgaa 240
|||||
Db 143990 TTCAGCCAGACCTTTCACTCAAGCCCTTGCAAAATATTATTGCTAGAAAATGAA 143931
QY 241 aatctaggagctatccaga 260
|||||
Db 143930 AATCTTAGGGTAAGTAAGA 143911

RESULT 15
AP001034 167934 bp DNA linear PRI 06-NOV-2001
LOCUS AP001034 Homo sapiens genomic DNA, chromosome 11q clone:RP11-729B4, complete
DEFINITION Homo sapiens genomic DNA, chromosome 11q clone:RP11-729B4, complete
AC090401
AC090401.5 GI:16751479
VERSION HTG.
KEYWORDS Homo sapiens DNA, clone:RP11-729B4.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 167934)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
DIRECT SUBMISSION
Submitted (06-JAN-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehirocho,Tsukumi-ku, Yokohama, Kanagawa 230-0845, Japan
(E-mail:hattori@gsr.riken.go.jp, URL:http://ngp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
On Nov 5, 2001 this sequence version replaced gi:11176992.
COMMENT Location/Qualifiers
FEATURES
source
1. 167934
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-729B4"
BASE COUNT 56061 a 33108 c 32009 g 46756 t
ORIGIN

Query Match 32.9%; Score 250.4; DB 9; Length 167934;

```

Best Local Similarity 97.7%; Pred. No. 1.le-46;
Matches 254; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 ttccagtgctccaggcagccctcagcacagaaagaaacatggtcttagactgaagtaccaa 60
Db 22624 TTCCAGTGCTCCAGGCAGCCTCAGCACAGAAAGAAAGACATGGTCTAGACTGAAGTACCAA 22683

Qy 61 ctaaatcatctcctttcaaatattaccgcagacaccatcatggattccaagcaccgcacacag 120
Db 22684 CTAAATCATCTCCTTTCAAAATTATCACCGACACACCATCATGGATTCAAGCACCGCACACAG 22743

Qy 121 tccgggttttctggtattttctccagaabaatcactgcttcagaatatgagtcacagaaact 180
Db 22744 TCCGGTGTGTTCTGGGTATTTCTCCAGAAATCAGTCTTCCAGAAATGATGATCCACAGAACT 22803

Qy 181 ttccagccacgaccccttttcaactcaagcccttgcacaaattatttgcagaaaaatgaa 240
Db 22804 TTCAGCCACGACCTTTTCAACTCAAGCCCTTGCAAAATTTATTGCTAGAAAAATGAA 22863

Qy 241 aatcttagggactatccaga 260
Db 22864 AATCTTAGGGGTAAGTAAGA 22883

Search completed: August 30, 2002, 02:46:47
Job time: 7480 sec

QY	350	tcaggatacattctggggctctgttttgctcatttaattctggagcctctcctaattgca	409
DB	388	TCAGATATCCATTCTGGGGCTCTGTTTGTTCATTAAATTCGGAGCCTTCCTAAATTCGA	329
QY	410	gtgaaagaaaaaacacagaaaactctgataataattgagccgcaataatgaattttcttagt	469
DB	328	GTGAAAGAAAAACACAGAAACTCTGATAATAATTAGAGCCGAATAATGAATTTCTTAGT	269
QY	470	gccttgagagaataagctggaatacttctcctcaatttgggttctcatctcagatacaaac	529
DB	268	GCCTTGGGAGCAATAGCTGGAAATCATTTCTCCACATTTGGTTTCATCCTAGATCAAAAC	209
QY	530	tacatttgggttattctcaccaaaaatgctcagtgtaagcgtcttactgtccctgtctctt	589
DB	208	TACATTGTGGTTATCTCCACCAAAATAGTCAGTGAAGCGTGTACTGTCTGTCTCTTG	149
QY	590	gaaatttgattacattgatgacttcacgaattattgaattattcattctctgccttcc	649
DB	148	GGAAATTTTGATTACATTGATGACATTTTCAGACATTTATGAATTTATTCATTTCTGCCTTC	89
QY	650	tcaatttgggtgcactcagagattgtgattgtgaacaattgttggtagctagcactg	709
DB	88	TCAATTTGGGTGCCACTCAGAGATGTGATGTGAACAATGTGTGTGACTAGCCTG	29
QY	710	tgagaataaagatgtgttaaatctcaa	737
DB	28	TGAGATAAGATGTGTTAAAAATAAAAA	1

RESULT 4

AA416972/c

LOCUS

DEFINITION

3' similar to SW:CD20_HUMAN P11836 B-LYMPHOCYTE ANTIGEN CD20 ; , mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

387 bp mRNA linear EST 09-NOV-1997

z94h05.sl Soares_testis_NHT Homo sapiens cDNA clone IMAGE:730041

Human.

AA416972

AA416972.1 GI:2077080

EST.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 387)

Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, X., Wyllie, T., Waterston, R., and Wilson, R.

WashU-NCI human EST Project

Unpublished (1997)

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LML ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 656 Std Error: 0.00

Seq primer: -4lm13 fwd. ET from Amersham.

Location/Qualifiers

1. .387

/organism="Homo sapiens"

/db_xref="GDB:5926570"

/db_xref="taxon:9606"

/clone="IMAGE:730041"

/clone_lib="Soares_testis_NHT"

/sex="male"

/lab_host="DH10B"

/note=Vector: p7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories , Inc., and primed with a Not I - oligo(dT) primer [5' ,

.
